THE APPLICATION OF ITERATIVE, DECOMPOSITION AND COMPARISON METHODS TO DUAL EXTREMUM PRINCIPLES

Thesis submitted for the degree of Doctor of Philosophy

at the University of Keele

by

PATRICIA BAILEY B.A.

Department of Mathematics
University of Keele
1984

DECLARATION

The material in this thesis is claimed as original except where explicity stated otherwise. This thesis has not been submitted previously for a higher degree of this or any other University.

ACKNOWLEDGEMENTS

I am extremely grateful to my supervisor, Dr Peter Smith, for the wonderful way he has guided and helped me during the past $3\frac{1}{2}$ years.

I would like to thank the Science and Engineering Research Council for their financial assistance.

ABSTRACT

This thesis is concerned with dual extremum principles and their applications. Dual extremum principles provide bounds to a functional associated with a given equation, and in some cases this functional is a measure of a physical quantity. These bounds can sometimes also provide information about the solution of the equation.

Chapter I is a survey of dual extremum principles from the classical maximum and minimum principles. Three important applications are summarised in this section.

Chapter II, on general principles, forms the basis of the thesis and includes definitions and theorems necessary for the following chapters. Sections on differential and integral operators, and convergence of iterative schemes, ends the chapter.

Chapter III develops optimising iterative schemes for dual extremum principles for linear problems. Convergence of the iterative schemes is considered and two examples complete the chapter.

In Chapter IV, we look at the decomposition of saddle functionals into two saddle functionals with a view to obtaining improved bounds. The method is applied to a particular functional, and conditions are found which ensure that the decomposition bounds are sharper than the classical bounds. The combination of iterative methods and decomposition dual extremum principles is considered, and conditions for convergence are found. The chapter ends with four examples.

Chapter V deals with comparison functionals, a method which involves finding simpler saddle functionals to approximate the saddle functional for which bounds are required. The theory is applied to the same

particular functional as that used in chapter IV, and conditions are developed which ensure that the comparison bounds are sharper than the classical bounds. The combination of comparison bounds and iterative methods is included, and conditions for convergence are found. Two examples end the chapter.

Chapter VI is a short chapter which looks at dual extremum principles for functionals which are convex/concave but not saddle over the whole of their domain. Three applications of the theory are included.

CONTENTS

CHAPTER V : COMPARISON FUNCTIONALS

CHAPTER VI : CONVEX/CONCAVE FUNCTIONALS

APPENDICES

NOTATION

Symbol	Meaning	Introduced in Section
$\in$	Belongs to	I.3
$[a, b]$	Closed interval consisting of the real numbers x such that $a \leq x \leq b$	I.3
$D(A)$	Domain of A	I.7
$\forall x$	For all x	I.6
∇	Gradient of a functional	I.3
I	Identity operator	II.4
$\langle u, v \rangle$	Inner product of u and v	I.2
$\langle u, v \rangle_D$	Inner product of u and v on the space D	I.4
∇^2	Laplacian operator divgrad	I.3
max	Maximum	I.3
min	Minimum	I.3
$\lvert a \rvert$	Modulus of a, where a is a real number	I.3
$\lVert a \rVert$	Norm of a, where a is an element of a vector space	I.3
$]a, b[$	Open interval	I.8
$\mathbb{R}$	Set of real numbers	I.9
$\subseteq$ $[\subset]$	Subset of (strict subset of)	I.7
sup	Supremum	II.4
O_n	Terms of order n	II.6

INTRODUCTION

This thesis looks at how iterative, decomposition and comparison methods can be used to improve the bounds obtained using dual extremum principles. Dual extremum principles provide a method for obtaining bounds to a functional associated with a given problem; for example, if D is an operator, linear or non-linear, f is a given function and ϕ is the solution of the equation $D\phi = f$, then by using dual extremum principles we can obtain bounds to the functional $\langle \phi, f \rangle$. This quantity is sometimes a useful number in its own right; even if it is not, it is sometimes possible to use the bounds to obtain information about the exact solution ϕ of $D\phi = f$.

The thesis starts with a survey. This traces the route to dual extremum principles from the maximum and minimum principles of Mikhlin and other authors and shows, in detail, how the concept of a saddle functional was crucial to the development of dual extremum principles; one section shows how the development of dual extremum principles over the years has influenced the work carried out on the magnetohydrodynamic pipe flow problem. Three important applications, Comparison Operators, Bivariational Bounds and Iterative Methods are summarised in this section.

Chapter II sets out the general principles necessary for the main part of this thesis. The chapter starts off with basic concepts such as vector space, inner product, operators, functionals and their derivatives; then convexity and concavity, saddle functionals and variational principles are considered. These lead into the dual extremum principles on which the thesis is based. The conditions which guarantee uniqueness are given next, and then several sections which look at the topics of differential and integral operators, convergence for bounded and unbounded operators, and cobweb iterative schemes. Many examples are interspersed throughout the chapter.

In chapter III, an optimising iterative method is developed for the general quadratic functional given by the equation $L(\phi, \psi) = \langle\phi, A\psi\rangle + \frac{1}{2}\langle\phi, B\phi\rangle - \frac{1}{2}\langle\psi, C\psi\rangle + \langle\phi, f\rangle + \langle\psi, g\rangle$, where A is a linear operator with an adjoint A^x and B and C are linear, symmetric, positive definate operators. The chapter starts by giving the classical dual extremum principles and cobweb iterative schemes for this functional; the next five sections are concerned with the development of the optimising iterative scheme. Section III.9 discusses convergence, and the last two sections considers two applications of the iterative methods.

Chapter IV looks at the decomposition of saddle functionals into two similarly orientated saddle functionals with the view to obtaining better bounds than those obtained with classical dual extremum principles. Section IV.2 sets out the basic theory, and the next section applies the theory to the general quadratic functional whose equation is given above. Conditions are found which ensure that the decomposition bounds are better than the classical bounds, and an example ends the section. Sections IV.4 and IV.5 look at the combination of iterative methods and decomposition dual extremum principles; the application of iterative methods to classical dual extremum principles is given for comparison, and conditions which guarantee convergence are developed. An example is also included which compares the conditions necessary for convergence using the various schemes. The last section, IV.6 applies the Decomposition of Functionals Theorem to four particular examples.

Chapter V is concerned with Comparison Functionals. This method involves finding bounds to the stationary value of $L(\phi, \psi)$ by using other, simpler functionals. The basic theory is set out in section V.2, and section V.3 applies two of the theorems from section V.2 to the usual quadratic functional, and finds conditions which ensure that the comparison bounds are better than the classical bounds. In section V.4, cobweb iterative schemes are applied to the comparison bounds obtained in section V.3, and conditions for convergence are found, and it is shown that operators exist which satisfy the convergence conditions. A brief note concerning iterative schemes for which convergence cannot be shown ends the section. Section V.5 looks at two applications of the methods developed in the chapter.

The final chapter, Chapter VI, is a short chapter which looks at bounds for functionals $L(\phi,\psi)$ which are not saddle over the whole of the domain of $L(\phi,\psi)$. Usually we obtain bounds for $L(\phi,\psi)$ which is convex in a set U_1, for all $\psi \in V_1$ and concave in a set V_1 for all $\phi \in U_1$. In this chapter we consider functionals $L(\phi,\psi)$ which are convex in a set U_2 for all $\psi \in V_1$ and concave in a set V_2 for all $\phi \in U_1$, where $U_1 \subseteq U_2$ and $V_1 \subseteq V_2$. Section VI.2 gives the basic theory and section VI.3 deals with applications.

CHAPTER I

I.1 Introduction

This chapter reviews the literature on Dual Extremum Principles. The number of papers published on this and related topics during the last twenty-five years is immense, and an exhaustive review would fill a large book. This survey will therefore concentrate mainly on the basic theory, with brief details of some of the problems to which the theory has been applied.

In Sections I.2 to I.4, the survey traces out the development from the classical maximum and minimum principles, through complementary variational principles to dual extremum principles; in Section I.5 it is shown how the dual extremum principles can be derived from a single functional.

Section I.6 looks at how the development of dual extremum principles from the classical maximum and minimum principles has influenced the work carried out on one particular problem, the magnetohydrodynamic pipe flow problem.

Sections I.7 to I.9 consider three important techniques used to extend dual extremum principles; these are Comparison Operators, Bivariational Bounds and Iterative Methods. The Survey ends by giving one example in which a well known result has been re-derived using one of these techniques.

CHAPTER I

I.2 Classical Extremum Principles

Dual extremum principles were developed from classical extremum principles, the basic ideas of which are summarised below; more details can be found in Mikhlin's book of 1964 (44), which contains an extensive account of the classical theory. The following summary comes from section 11 of (44).

If we have an operator equation

$$Au_e = f \tag{I.2.1}$$

where A is a positive-definite symmetric operator, f is a given function and u_e is the unknown solution of the equation, then the Minimum Functional Theorem states that the functional

$$F(u) = \langle u, Au\rangle - 2\langle u, f\rangle \tag{I.2.2}$$

has a minimum if u_e is the unique solution of equation (I.2.1).
Substituting equation (I.2.1) into (I.2.2) gives

$$\begin{aligned} F(u_e) &= \langle u_e, Au_e\rangle - 2\langle u_e, Au_e\rangle \\ &= -\langle u_e, Au_e\rangle = -\langle u_e, f\rangle \end{aligned} \tag{I.2.3}$$

$$\begin{aligned} \text{Now } \langle u, f\rangle - \langle u_e, f\rangle &= \langle u - u_e, Au_e\rangle \\ &= -\langle u - u_e, A(u - u_e)\rangle + \langle Au - f, u\rangle \end{aligned}$$

Since A is positive-definite,

$$\langle u, f\rangle - \langle u_e, f\rangle \leq \langle Au - f, u\rangle$$

hence $-\langle u_e, f\rangle \leq \langle Au - 2f, u\rangle$

or $F(u_e) \leq \langle Au - 2f, u\rangle$ for any u (I.2.4)

In some problems of mathematical physics, the magnitude of the functional F(u) is proportional to the potential energy of the system. The minimum functional theorem is then equivalent to the principle of minimum potential energy. For this reason, the quantity $\langle u, Au\rangle$ is often called the ENERGY of the system and the minimum principle is called the ENERGY PRINCIPLE.

CHAPTER I

The method involves finding a function which minimizes $F(u)$. It is possible that a different functional $G(u)$ exists which has a maximum when u is the solution of $Au_e = f$, thus giving both upper and lower bounds to the energy.

Obtaining a function u which minimizes $F(u)$ is carried out by finding suitable trial functions and substituting them in turn into equation (I.2.2); the minimum value of $F(u)$ can then be taken as a good approximation to $F(u_e)$. In some problems, the trial functions are arbitrary, for example when A is an integral operator. In problems in which A is a differential operator, the trial functions must satisfy given boundary conditions, which are generally required to ensure that A is positive-definite. Suitable trial functions can sometimes be found systematically: if a sequence $u_1, u_2, u_3 \ldots u_n$ can be found, and $\lim_{n\to\infty} F(u_n) = F(u_e)$, then the sequence could be a minimizing sequence and may converge to the unique solution u_e of $Au_e = f$. Methods for constructing the minimizing sequence include the Ritz method and the method of steepest descent.

The minimum functional theorem is applied to many examples in (44); these include

(a) The second order differential equation

$$-\frac{d}{dx}\left(p(x)\frac{du}{dx}\right) + r(x)\,u = f(x), \quad \alpha_1 u'(a) - \beta_1 u(a) = 0,$$

$$\alpha_2 u'(b) + \beta_2 u(b) = 0, \quad a \leq x \leq b$$

(b) The bending of a beam: $\frac{d^2}{dx^2}\left[EI(x)\frac{d^2w}{dx^2}\right] + Kw = q(x)$,

$$0 \leq x \leq 1; \quad w(o) = w(1) = w'(o) = w'(1) = 0$$

(c) Poisson's problem $-\nabla^2\phi = f(p)$ on V, $\phi = 0$ on S , and other related problems.

(d) Equations in elasticity theory.

(e) Bending of thin plates.

(f) Finding bounds in the minimum eigenvalue λ_0 of an operator A where $Au = \lambda u$.

I.3 Complementary Variational Principles

Complementary variational principles arise when a problem can be formulated variationally in two different but related ways, and both upper and lower bounds can be obtained to some quantity related to the problem. The method consists of calculating the values of these complementary functionals when suitable trial functions are taken. The quantity bounded may be of some physical significance (for example the energy of the system as considered in the previous section); and in situations where this does not apply, the closeness of the upper and lower bounds may give some indication of how close the trial functions are to the solution of the original problem.

Some results published prior to 1964, when the term 'complementary variational principles' was first introduced, are now recognised as falling into the sphere of complementary variational principles: for example, Fujita in 1955 (38) gave upper and lower bounds to the quantity $\|Tu_0\|^2$, where $T^x Tu_0 = f$. However, the impetus for the rapid development of the subject came with a paper by Noble in 1964 (48). A paper by Rall in 1966 (52) put the ideas of Noble into a functional analytic framework, and this was rapidly followed by a large number of papers which considered various applications of complementary variational principles, mainly by A M Arthurs, P D Robinson and N Anderson. Many of the results obtained were collected in a book published by Arthurs in 1970 (7). This book included many varied examples; papers published afterwards expanded on some of these examples.

Rall's account is based on the canonical equations

$$T\phi - \frac{\partial W}{\partial u}(u, \phi) = 0, \qquad \text{(I.3.1)}$$

$$T^x u - \frac{\partial W}{\partial \phi}(u, \phi) = 0 \qquad \text{(I.3.2)}$$

(Note that we are using '$\frac{\partial}{\partial u}$' and '$\frac{\partial}{\partial \phi}$' rather than the now more usual '∇u' and '$\nabla \phi$' to signify the functional derivations, as this was the notation used in Rall's paper.)

In equations (I.3.1) and (I.3.2), T is a linear operator with an adjoint T^x such that

$$\langle u, T\phi \rangle = \langle \phi, T^x u \rangle \qquad (I.3.3)$$

and $W(u, \phi)$ is a twice differentiable functional.

Equations (I.3.1) and (I.3.2) are said to arise from a variational principle if $F(u, \phi) = f'(u, \phi)$, where $f(u, \phi)$ is a functional and

$$F(u, \phi) = \begin{pmatrix} T\phi - \frac{\partial W}{\partial u} \\ T^x u - \frac{\partial W}{\partial \phi} \end{pmatrix} \qquad (I.3.4)$$

By integration, $f(u, \phi) = \langle u, T\phi \rangle - W(u, \phi) = f_1$ (I.3.5)

$$\equiv \langle \phi, T^x u \rangle - W(u, \phi) = f_2 \qquad (I.3.6)$$

A solution (u_e, ϕ_e) of the system given by equations (I.3.1) and (I.3.2) is a stationary point of $f(u, \phi)$. If equation (I.3.1) arises from a maximum principle for $f_1(u)$ for fixed ϕ and equation (I.3.2) arises from a minimum principle for $f_2(\phi)$ for fixed u (or vice-versa), then equations (I.3.1) and (I.3.2) are said to arise from complementary variational principles. In this case

$$f(u_e, \phi_e) = \max_u \min_\phi f(u, \phi) = \min_\phi \max_u f(u, \phi) \qquad (I.3.7)$$

and the fixed point (u_e, ϕ_e) is called the MINIMAX point of $f(u, \phi)$; it is then possible to obtain both upper and lower bounds to $f(u_e, \phi_e)$.

CHAPTER I

The main results of this paper are given in the following theorem:

Theorem (I.3.1)

If (u_α, ϕ_α) satisfies $T^x u = \frac{\partial W}{\partial \phi}$ and (u_β, ϕ_β) satisfies $T\phi = \frac{\partial W}{\partial u}$, then

$$f_1(u_\beta, \phi_\beta) \leq f(u_e, \phi_e) \leq f_2(u_\alpha, \phi_\alpha) \qquad \text{(I.3.8)}$$

where f_1 and f_2 are given by equations (I.3.5) and (I.3.6) and (u_e, ϕ_e) satisfies both $T^x u = \frac{\partial W}{\partial \phi}$ and $T\phi = \frac{\partial W}{\partial u}$.

A sufficient condition for (u_e, ϕ_e) to be a minimax point, and for (I.3.1) and (I.3.2) to arise from complementary variational principles is that $\frac{\partial^2 W}{\partial u^2} \geq 0$ and $\frac{\partial^2 W}{\partial \phi^2} \leq 0$.

these bounds are local bounds.

Arthur's monograph of 1970 expanded on the ideas of Noble and Rall. After a chapter detailing the basic theory of variational and complementary variational principles, Arthurs considered a class of operators which satisfy the equation

$$\langle u, T\phi \rangle = \langle \phi, T^x u \rangle + (S(u, \phi)) \qquad \text{(I.3.9)}$$

where T^x is the formal adjoint of the linear operator T and $(S(u,\phi))$, the conjunct of u and ϕ, denotes boundary terms. Examples of T, T^x and $(S(u,\phi))$ given in the book are:

(i) $T = d/dx$, $T^x = -d/dx$, $(S(u,\phi)) = \Big[u(x)\,\phi(x)\Big]_a^b$

(ii) $T = \text{grad}$, $T^x = -\text{div}$ $(S(u,\phi)) = \int_{\partial V} \underline{u}\cdot\underline{n}\,\phi\, dB$

(iii) $T = T^x = \text{curl}$, $(S(u,\phi)) = \int_{\partial V} u\cdot(\underline{n}\wedge\phi)\, dB$

(iv) Integral operators: $T\phi(x) = \int_a^b k(x,y)\phi(y)\,dy$, $T^x u(x) = \int_a^b K^x(x,y)\, u(y)\, dy$, where $k(y,x) = K^x(x,y)$ is any continuous function of x and y; $(S(u,\phi)) = 0$

(v) T is a m x n matrix; then $T^x = T^t$, the transpose of T and $(S(u,\phi)) = 0$

CHAPTER I

The theory from the first chapter is then extended to this class of operators, and is finally recast in an abstract form which results in a theorem equivalent to theorem (I.3.1) above, although with different notation; boundary terms are included in the upper and lower bounds.

The book continues by applying the theory to the problem

$$(T^x T + Q)\phi = f \text{ in } V, \quad \sigma_r(\phi - \phi_s) = 0 \text{ on } \partial V \qquad \text{(I.3.10)}$$

where T and T^x are linear operators satisfying equation (I.3.9), $(S(u,\phi)) = (u, \sigma_r \phi)$ and Q is a positive, symmetric operator. After summarising the complementary variational principles for the problem, the Rayleigh-Ritz method for optimising the upper and lower bounds is discussed, and then the theory is applied to a variety of problems, for some of which numerical results are given. These include:

(a) The eigenvalue problem (Example (a) in section I.2)

(b) The Dirichlet problem: $\nabla^2\phi = 0$ in V, $\phi = \phi_s$ on V.

(c) The Electrostatic field equation, $\nabla^2\phi = -4\pi p$ in V, $\phi = \phi_s$ on ∂V.

(d) The Magnetostatic analogue of (c), curl curl $\underline{\phi} = 4\pi \underline{j}$ in V, $\underline{\phi} = \underline{\phi}_s$ on ∂V.

(e) A problem in diffusion, $(-\nabla^2 + \kappa^2)\phi = K$ in V, $\phi = 0$ on ∂V.

(f) The Milne problem on Neutron transport,
$\phi(x) = f(x) + K\phi(x)$, where $f(x) = (c/2v)\,(E_1(x) - e^{vx} E_1\{(1 + v)x\})$,
$K\phi(x) = \frac{c}{2}\int_0^\infty E_1(|x - t|)\,\phi(t)\,dt$, $E_1(x) = \left[\int_1^\infty e^{-xt}\,dt\right]/t$,
v is determined by $2v = c \ln\left[(1 + v)/(1 - v)\right]$ and $0 < c < 1$.

(g) The Kirkwood-Riseman integral equation
$\phi(x) = f(x) + \lambda\int_{-1}^{1} |x - t|^{-\alpha}\,\phi(t)\,dt \quad 0 < \alpha < 1$

(h) Problems from quantum theory - perturbation theory and potential theory using both differential and integral equation approaches.

The last chapter considers a class of non-linear problems given by the equations $T^x T\phi = F(\phi)$ in V, $\sigma_r(\phi - \phi_s) = 0$ on ∂V (I.3.11)

where T and T^x are linear operators satisfying equation (I.3.9), with $(S(u,\phi)) = (u, \sigma_r\phi)$ and $F(\phi)$ is a given non-linear operator.

After summarising the complementary variational principles for the problem, applications are discussed which include:

(i) The Liouville equation $\nabla^2\phi = ce^{\phi}$ in V, $\phi = \phi_s$ on ∂V.

(j) The Poisson-Boltzmann equation: $\phi'' = 2\sinh\phi$, $x \in [0, a]$; $\phi(0) = \phi_1$, $\phi(a) = \phi_2$.

(k) The Thomas-Fermi equation: $x^{\frac{1}{2}}\phi'' = \phi^{\frac{3}{2}}$, $0 \leq x < \infty$, with $\phi(0) = 1$, $\lim\limits_{x\to\infty} \phi(x) = 0$, $\lim\limits_{x\to\infty} x\phi'(x) = 0$.

(l) A non-linear equation from communication theory, $\phi K\phi = 1$, where $K\phi(x) = \int_0^{\frac{\pi}{2}} \left\{[\sin(x - y)] \Big/ [\pi(x - y)]\right\} \phi(y)\, dy$, $0 \leq x \leq \frac{\pi}{2}$

(m) Equations containing matrix operators, arising from problems in electrical networks.

(n) A problem modelling the steady flow of a non-viscous compressible fluid, with equations

$\operatorname{div}(p\,\underline{q}) = 0$, $(\underline{q}\,.\,\mathrm{grad})\,\underline{q} = -\,\mathrm{grad}\,p/p$

As before, numerical results are given in some cases.

Many of the papers published after this book expanded on the problems listed above, usually by considering different boundary conditions or new numerical results. New problems treated in papers published around the same time included:

(o) $\phi^{iv} = F(\phi)$, $\phi(0) = \phi(1) = \phi'(0) = \phi'(1) = 0$ (Anderson and Arthurs, 1970, (2)).

(p) $L\phi = \sum\limits_{k=0}^{m} (-1)^k D^k\{p_k(x)\, D^k\phi\} = f(x)$, $a < x < b$, where D denotes differentiation, $p_0 > 0$, $p_m > 0$, $p_k \geq 0$, $(k = 1,2,\ldots m - 1)$; and boundary conditions are $D^k\phi(a) = D^k\phi(b) = 0$, $k = (0,1,\ldots m - 1)$, $m \geq 1$ (Arthurs and Coles, 1972, (11)).

(q) $\phi'' = (1 + r\phi)^{-\frac{1}{2}} - e^{-\phi}$, $\phi(0) = 0$, $\phi'(0) = X$ (Arthurs and Anderson, 1968, (10)).

I.4 Dual Extremum Principles

The next step forward was provided by Sewell in his 1969 paper (57). As we saw in the previous section a sufficient condition for a solution (u_e, ϕ_e) of the pair of equations $T\phi = \partial W/\partial u$ and $T^x u = \partial W/\partial \phi$ to be a minimax point is $\partial^2 W/\partial u^2 \geq 0$ and $\partial^2 W/\partial \phi^2 \leq 0$ everywhere (or vice-versa). However, there is no guarantee that (u_e, ϕ_e) is unique.

In his 1969 paper, Sewell recognised that convexity forms the basis of extremum principles and uniqueness theorems. The paper showed that if $W(u,\phi)$ is a saddle functional (that is, convex in u for fixed ϕ and concave in ϕ for fixed u, or vice-versa) then complementary bounds exist. Moreover, if $W(u,\phi)$ is a strict saddle functional then (u_e, ϕ_e) is unique.

Three review papers then followed which exploited the concept of saddle functionals. Complementary variational principles were renamed 'Dual Extremum Principles' by Noble and Sewell in their 1972 paper (50).

The first review paper, by Robinson in 1971 (53) starts by giving a short history of the development leading to dual extremum principles. Variational theory is then considered, as are complementary variational principles, in a manner similar to that of Arthurs (7) ; abstract formulation including boundary terms, is included. Four basic types of boundary conditions are listed, and in each case it is stated if the boundary condition is essential (must be satisfied by the trial functions) or natural (need not be satisfied by the trial functions) for each bound. Robinson next proves the sufficiency conditions for dual extremum principles, and for the uniqueness of the stationary point (u_e, ϕ_e) ; this is followed by examples of choices of the operators T and T^x, with the appropriate boundary terms.

The next chapter considers the class of problems described by the equations

$$T^x T\phi + f(\phi) = 0 \text{ in } V, \quad \phi = \alpha \text{ on } \partial V \qquad (I.4.1)$$

The functional $W(u,\phi)$ is shown to take the form

$$W(u,\phi) = \tfrac{1}{2}\langle u, u\rangle - \langle 1, F(\phi)\rangle \qquad (I.4.2)$$

where $f(\phi) = F'(\phi)$ and $f'(\phi) > 0$ to ensure that $W(u,\phi)$ is a strict convex/concave saddle functional. After giving conditions under which it is not always necessary to know T and T^x individually, the iterative scheme

$$T^x T\phi_{n+1} + f(\phi_n) = 0, \quad n = 0,1,2,\ldots \qquad (I.4.3)$$

is considered. Error bounds for an approximate solution are also dealt with.

A chapter follows giving dual extremum principles for the Hammerstein integral equation,

$\phi + K(e(\phi)) = 0$ in V, where

$K(e\phi(s)) = \int_a^b k(s,t)\, e(\phi(t))\, dt, \quad a \leq s \leq t$

The connection between the bounds for this equation and those for its associated differential equation is briefly considered.

The last chapter consists of the dual extremum principles for several non-linear problems; in most cases numerical results are given. Problems treated are:

(a) Poisson-Boltzmann equation.

(b) Thomas-Fermi equation.

(c) Foppl-Hencky equation, $-\frac{d}{dx}\left(x^3 \frac{d\phi}{dx}\right) - \frac{2x^3}{\phi^2} = 0, \quad 0 \leq x \leq 1,$

$\phi'(0) = 0, \quad \phi(1) = \lambda > 0$

(d) The non-linear equation from communication theory.

(e) A non-linear Kirkwood integral equation.

$\psi(x) = \frac{\lambda}{4}\int_{-\infty}^{\infty} k(x - x')\, x'\,(g(x') - 1)\, dx' \quad (\lambda > 0)$

(f) A transportation network problem.

(g) Compressible fluid flow.

(h) Heat loss from a cell - a linear equation with a non-linear boundary condition: $\nabla^2\phi = 0$ in V, $\phi = 0$ on ∂V_1, $\underline{n} . \nabla\phi = -\lambda\phi^4$ on ∂V_2 $(\lambda > 0)$, $\underline{n} . \nabla\phi = 0$ on ∂V_3.

Problems (a), (b), (d), (f) and (g) were detailed in the last section. The paper ends with an appendix on convex functions and functionals.

The second review paper, by Noble and Sewell in 1972 (50), presents an account of the theory of dual extremum principles, including convexity and saddle functionals. An appendix on functional analysis covering topics such as inner product, operators and their adjoints, functional gradients and convexity, ends the paper.

The approach of this paper is the reverse of the usual one, in which a problem is given and a pair of canonical equations leading to dual extremum principles are derived. The paper considers several groups of canonical equations and derives, by an automatic procedure, dual extremum principles for the solutions of the problems defined by each group.
After the introduction, the paper lists two groups, each containing four sets of governing conditions. Most of these sets consist of inequalities rather than equations. The governing conditions analogous to those in Arthurs (7) and Robinson (53) are

$$T^x u = \partial X/\partial x \quad (\alpha), \quad Tx = \partial X/\partial u \quad (\beta) \qquad (I.4.4)$$

It is noted that only in those problems consisting solely of equations can we expect the extrema to be stationary.

The next section explains the idea of a Legendre transformation, by means of which the conditions in one group can be transformed into those of another. However, as in this survey we are only concerned with problems whose defining equations are given by (I.4.4), Legendre transformations will not be pursued here.

CHAPTER I

Sections 4 and 5 set out the theory of convexity and saddle functions; both sections include several examples. The next section looks at the relationships between convex and saddle functions, with reference to Legendre transformations. Section 7 considers uniqueness. It is proved that if (x_e, u_e) is a solution of the problem given by equation (I.4.4), this solution is unique provided $X(x, u)$ is a strict saddle functional. It is noted that if $X(x, u)$ is only partly strict, say, for example, strictly convex in x and weakly concave in u, then x_e is unique but nothing can be inferred about the uniqueness of u_e.

Section 8 sets out the dual extremum principles for each group of conditions, and gives proofs. The proof depends on the saddle property of $X(x, u)$; it is noted that a weak saddle is sufficient for the existance of dual extremum principles. The bounds relating to (I.4.4) are:

$$K(\beta) \leq \max_{\beta} K(\beta) = \min_{\alpha} J(\alpha) \leq J(\alpha) \qquad \text{(I.4.5)}$$

where $K(\beta) = \langle u, \partial X/\partial u\rangle - X(x, u)$ (I.4.6)

and $J(\alpha) = \langle x, \partial X/\partial x\rangle - X(x, u)$ (I.4.7)

Functions to be inserted into $K(\beta)$ satisfy equation $(I.4.4)_\beta$, those to be inserted into $J(\alpha)$ satisfy equation $(I.4.4)_\alpha$, and the solution (x_e, u_e) giving $\max_{\beta} K(\beta) = \min_{\alpha} J(\alpha)$ satisfies both $(I.4.4)\alpha$ and β.

Section 9 concerns boundary value problems. In (7) boundary terms were not included in the operator T or its adjoint, but were placed in an extra term called the conjunct, denoted by $(S(u, x))$. Thus we have

$$\langle u, Tx\rangle = \langle x, T^{x} u\rangle + (S(u, x)) \qquad \text{(I.4.8)}$$

In this paper it is shown that boundary terms can be incorporated into T and X, so that they are included in the canonical equations. If we let

$T = D$ in V, N on V (I.4.9)

and $X = H$ in V, B on V (I.4.10)

with adjointness given by

$$\langle u, Dx\rangle_V + \langle u, Nx\rangle_V = \langle x, D^x u\rangle_V + \langle x, N^x u\rangle_V \qquad (I.4.11)$$

then equation (I.4.4) can be written

$$\begin{pmatrix} D^x u = \partial H/\partial x \text{ in } V \\ N^x u = \partial B/\partial x \text{ on } \partial V \end{pmatrix}_\alpha , \qquad \begin{pmatrix} Dx = \partial H/\partial u \text{ in } V \\ Nx = \partial B/\partial u \text{ on } \partial V \end{pmatrix}_\beta \qquad (I.4.12)$$

It is stated that $X(x, u)$ $(= H(x, u) + B(x, u))$ will be a strict saddle functional provided that both $H(x, u)$ and $B(x, u)$ are saddle functionals with the same orientation, and at least one of them is a strict saddle functional. In this section of the paper, a procedure is given in which the boundary terms in a problem for which we want to find dual extremum principles are used to find the form of $B(x, u)$. The idea is to specify $B(x, u)$ so that the equations $N^x u = \partial B/\partial x$ and $Nx = \partial B/\partial u$ result in the boundary terms of the given problem. Natural and essential conditions are discussed, in a similar manner to their discussion in Robinson (53), and the boundary terms to be included in $K(\beta)$ and $J(\alpha)$ are tabulated. The theory is applied to an ordinary differential equation and a partial differential equation.

The rest of the paper deals with applications. Problems covered are from the following areas:

(i) Finite mathematical programming; (j) Network theory; (k) Compressible fluid flow; (l) Elasticity; (m) Plasticity; (n) Optimisation and control theory; (o) Integral equations; (p) Analysis.

The last of these review papers, by Arthurs in 1973 (8) looks at dual extremum principles and error bounds for boundary value problems. In this paper, Arthurs extends his previous work on the theory of complementary variational principles (7) to take into account the concept of saddle functionals; the four basic types of boundary conditions for differential equations (as previously discussed by Robinson, (53)) are discussed in detail.

The dual extremum principles for the equation $T^x T\phi = f(\phi)$, where T is a differential operator, is then dealt with, with particular emphasis on the boundary conditions. Terms for the error bound $\Delta J = J(\alpha) - K(\beta)$ are derived in this part of the paper. Paper (8) ends with two illustrations of the theory, to

(q) $\phi''(x) = C\phi - p,\ 0 \leq x \leq 1$, with $\phi'(0) = 0, \quad \phi'(1) = -u\phi^4$

(r) $\nabla^2\phi = -1, \ -1 \leq x \leq 1, \ -1 \leq y \leq 1$, with $\phi(x, \pm 1) = 0$

and $\frac{\partial\phi}{\partial n}(\pm 1, y) = -\phi$

Following these three review papers, there were many papers on aspects of dual extremum principles. Some of these covered problems which had been dealt with previously, and were now considered in the light of the concept of saddle functionals; other papers looked at new examples, some with numerical results, including

(s) A non-linear heat transfer problem,

$$\phi''(x) = \varepsilon\phi^4 + \lambda^2\phi - \varepsilon\phi_e^4 - \lambda^2\phi_e, \quad 0 \leq x \leq 1,$$

$$\phi(0) = 1, \quad \phi'(1) = -\varepsilon v\left[\phi^4(1) - \phi_e^4\right]$$

Anderson and Arthurs, 1972 (3)

(t) After finding dual extremum principles for $(1 + yL)\phi = f$, where L is a non-negative symmetric linear operator and $y \in R$, series expansion trial functions are used to derive families of approximants for $\langle\phi, f\rangle$

Barnsley and Robinson, 1974 (17)

(u) A problem in the twisting of ring sectors,

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{3}{R - x}\frac{\partial\psi}{\partial x} + 2C = 0 \text{ in } V, \quad \psi = 0 \text{ on } \partial V$$

Arthurs and Duggan, 1975 (12)

(v) A problem arising from a model of the human eye:

$K \nabla^2 p_m + \text{grad}\, p_m \cdot \text{grad}\, K = 0$ in V, $p_m = a$ on B_1

$\frac{\partial p_m}{\partial n} = b$ on B_2

Anderson and Arthurs, 1976 (4)

(w) A linear initial value problem:

$Aq''(t) + Bq'(t) + Cq(t) = f, \quad t > 0; \quad q(0) = a, \quad q'(0) = b$

Arthurs and Jones, 1976 (14)

(x) Complex equations: $L(\psi) = f(\psi)$ where f is a complex function of the complex vector ψ

Anderson and Arthurs, 1976 (5)

(y) The connection between extremum principles and the hypercircle-geometrical approach to the question of solving problems is investigated, with reference to the problem

$\nabla^4\phi - \text{div}\,(M \nabla \phi) = f(x, y)$ in V, $\phi = 0$ on B,

$\frac{\partial \phi}{\partial n} = 0$ on B_1, $\frac{\partial \phi}{\partial n} = - b \nabla^2\phi$ on B_2, where $B = B_1 + B_2$

Arthurs and Duggan, 1977 (13)

CHAPTER I

I.5 Dual Extremum Principles generated from a single functional

The previous two sections have shown how, given a particular problem, dual extremum principles can be found by finding a linear operator T with an adjoint T^{*} and a saddle functional $X(x, u)$ such that the pair of equations $Tx = \partial X/\partial u$ and $T^{*}u = \partial X/\partial x$ together define the given problem.

In a paper published in 1973, (58), Sewell suggested a new framework for obtaining dual extremum principles. He showed that the canonical equations can be generated from a single functional, and the dual extremum principles are then obtained from this functional, assuming that it is saddle. The aim when faced with a problem for which dual extremum principles are required then becomes one of finding this saddle functional. The saddle functional leading to a particular problem may not be unique; an example illustrating this will be given after the following summary of the results. These results form the basis of this thesis and will be considered in more detail in the next chapter. As in section I.3, we are still using '$\frac{\partial}{\partial u}$' and '$\frac{\partial}{\partial x}$' for the partial derivations, as this is the notation used in the paper.

If we are given the problem defined by the equations

$$T^{*} u = \frac{\partial X}{\partial x} \quad \text{and} \quad Tx = \frac{\partial X}{\partial u} \tag{I.5.1}$$

we construct a functional

$$L(u, x) = \langle x, T^{*} u\rangle - X(u, x) \tag{I.5.2}$$

$$\equiv \langle u, Tx\rangle - X(u, x) \tag{I.5.3}$$

such that $\frac{\partial L}{\partial x} = T^{*} u - \frac{\partial X}{\partial x}$ (α) and $\frac{\partial L}{\partial u} = Tx - \frac{\partial X}{\partial u}$ (β) (I.5.4)

To obtain dual extremum principles, we require that $L(u,x)$ is a saddle functional, that is concave in x for fixed u and convex in u for fixed x. $L(u, x)$ will be a saddle functional if it satisfies the following inequality:

I.5 Dual Extremum Principles generated from a single functional

The previous two sections have shown how, given a particular problem, dual extremum principles can be found by finding a linear operator T with an adjoint T^x and a saddle functional $X(x, u)$ such that the pair of equations $Tx = \partial X/\partial u$ and $T^x u = \partial X/\partial x$ together define the given problem.

In a paper published in 1973, (58), Sewell suggested a new framework for obtaining dual extremum principles. He showed that the canonical equations can be generated from a single functional, and the dual extremum principles are then obtained from this functional, assuming that it is saddle. The aim when faced with a problem for which dual extremum principles are required then becomes one of finding this saddle functional. The saddle functional leading to a particular problem may not be unique; an example illustrating this will be given after the following summary of the results. These results form the basis of this thesis and will be considered in more detail in the next chapter. As in section I.3, we are still using '$\frac{\partial}{\partial u}$' and '$\frac{\partial}{\partial x}$' for the partial derivations, as this is the notation used in the paper.

If we are given the problem defined by the equations

$$T^x u = \frac{\partial X}{\partial x} \quad \text{and} \quad Tx = \frac{\partial X}{\partial u} \tag{I.5.1}$$

we construct a functional

$$L(u, x) = \langle x, T^x u\rangle - X(u, x) \tag{I.5.2}$$

$$\quad\quad = \langle u, Tx\rangle - X(u, x) \tag{I.5.3}$$

such that $\frac{\partial L}{\partial x} = T^x u - \frac{\partial X}{\partial x}$ (α) and $\frac{\partial L}{\partial u} = Tx - \frac{\partial X}{\partial u}$ (β) (I.5.4)

To obtain dual extremum principles, we require that $L(u,x)$ is a saddle functional, that is concave in x for fixed u and convex in u for fixed x. $L(u, x)$ will be a saddle functional if it satisfies the following inequality:

$$L(u_+, x_+) - L(u_-, x_-) - \left\langle x_+ - x_-, \left.\frac{\partial L}{\partial x}\right|_+ \right\rangle$$

$$- \left\langle u_+ - u_-, \left.\frac{\partial L}{\partial u}\right|_- \right\rangle \geq 0 \tag{I.5.5}$$

where (u_+, x_+) and (u_-, x_-) are any distinct pairs of points in the domain of $L(u, x)$. The saddle property is strict if the inequality in equation (I.5.5) is strict. The dual extremum principles then follow:

$$L(u_\beta, x_\beta) \leq L(u_e, x_e) \leq L(u_\alpha, x_\alpha) \tag{I.5.6}$$

where $\frac{\partial L}{\partial u}(u_\beta, x_\beta) = 0$, $\frac{\partial L}{\partial x}(u_\alpha, x_\alpha) = 0$ (I.5.7)

and (u_e, x_e) satisfies both $\frac{\partial L}{\partial u} = 0$ and $\frac{\partial L}{\partial x} = 0$. The stationary point (u_e, x_e) is unique provided that $L(u, x)$ is a strict saddle functional.

A linear problem has a functional which is quadratic in its variables. A broad class of quadratic saddle functionals are given by the equation

$$L(u, x) = \langle u, Tx\rangle + \tfrac{1}{2}\langle u, Bu\rangle - \tfrac{1}{2}\langle x, Ax\rangle$$
$$+ \langle x, a\rangle + \langle u, b\rangle \tag{I.5.8}$$

where A and B are non-negative symmetric operators and T has an adjoint T^x. Using equation (I.5.8) we can now show that a particular problem can arise from more than one saddle functional. If we let $T = A = I$, $B = K$ and $a = 0$ in equation (I.5.8), where K is any linear, symmetric non-negative operator, then $L(u, x)$ becomes

$$L(u, x) = \langle u, x\rangle + \tfrac{1}{2}\langle u, Ku\rangle - \tfrac{1}{2}\langle x, x\rangle + \langle u, b\rangle \tag{I.5.9}$$

The gradients of equation (I.5.9) are

$$\frac{\partial L}{\partial x} = u - x, \quad \frac{\partial L}{\partial u} = x + Ku + b \tag{I.5.10}$$

The stationary point (u_e, x_e) satisfies $\frac{\partial L}{\partial x} = 0$, $\frac{\partial L}{\partial u} = 0$, giving $u_e - x_e = 0$ and $x_e + Ku_e + b = 0$, or $(I + K)u_e + b = 0$.

On the other hand, if we let $T = \frac{1}{2}I$, $B = K + \frac{1}{2}I$, $A = \frac{1}{2}I$ and $a = 0$ in equation (I.5.8), where K is the same operator as before, then $L(u, x)$ becomes

$$L(u, x) = \tfrac{1}{2}\langle u, x\rangle + \tfrac{1}{2}\langle u, (K + \tfrac{1}{2}I)u\rangle - \tfrac{1}{2}\langle x, \tfrac{1}{2}x\rangle + \langle u, b\rangle \tag{I.5.11}$$

The gradients of equation (I.5.11) are

$$\frac{\partial L}{\partial x} = \tfrac{1}{2}u - \tfrac{1}{2}x \quad \text{and} \quad \frac{\partial L}{\partial u} = \tfrac{1}{2}x + (K + \tfrac{1}{2}I)u + b \tag{I.5.12}$$

The stationary point (u_e, x_e) satisfies $\frac{\partial L}{\partial x} = 0$ and $\frac{\partial L}{\partial u} = 0$, resulting in $\frac{1}{2}u_e - \frac{1}{2}x_e = 0$ and $\frac{1}{2}x_e + (K + \frac{1}{2}I)u_e + b = 0$ or $(K + I)u_e + b = 0$, as before.

The theory given in this paper is illustrated by applying it to problems in elasticity and plasticity. Dual extremum principles using the saddle functional given by equation (I.5.8) were discussed by Smith in 1976 (64), where the theory is applied to the problem in magnetohydrodynamic pipe flow described by the coupled equations

(a) $\nabla^2 w + M\,\partial b/\partial y = -1$ in D

$\nabla^2 b + M\,\partial w/\partial y = 0$ in D

$w = b = 0$ on ∂D

Examples dealt with in a 1979 paper by Sewell (59) include those mentioned above and also

(b) Small deflection of a cantilever beam, described by the equations

$M''(s) - w + Pu''(s) = 0$ for $0 \leq s \leq 1$, $u(0) = M(1) = 0$

(c) $\nabla^2\phi = F'(\phi)$ in V, $\phi = h$ on S

(d) $\nabla\phi = \partial P/\partial \underline{Q}$ in V, $\nabla \cdot \underline{Q} = 0$ in V, $\underline{n} \cdot \underline{Q} = C$ on S

There are other papers which use Sewell's new framework and apply it to problems previously considered. Many of these results were collected together in the second edition (1980) of Arthurs' monograph on Complementary Variational Principles (9) . This book follows a similar format to the 1970 edition (7) but replaces the local theory with the global treatment based on convexity and saddle functionals; these two crucial concepts are considered in detail. A section is also included which shows how functional derivatives can be obtained. The second chapter ends with the main dual extremum principles result given earlier in this section.

Chapter 3 covers the class of problems described by the equation $(T^x T + Q)\phi = f$ in V; various boundary conditions are considered. This chapter includes sections on the hypercircle and error estimates, and also comparison operators and bivariational bounds, both of which will be looked at in more detail in the following sections.

The next chapter gives details of linear applications. All of those detailed in (7) are included, see examples (a) to (h) in section I.3 , some with new numerical results; other problems include

(e) A problem in classical elasticity.

(f) A problem in heat transfer, $(\nabla^2 - \underline{K}.\nabla)\phi = f$ in V, $\phi = 0$ on ∂V, div $\underline{K} = 0$.

Chapter 5 concentrates on the class of non-linear problems given by the equations

$T\phi = f_1(u)$, $T^x u = f_2(\phi)$ in V,

$\sigma\phi = m(u)$ on ∂V_1, $\sigma^x u = n(\phi)$ on ∂V_2

After developing the dual extremum principles for this class of problems, they are applied to

$T^x T\phi = f(\phi)$

with suitable boundary terms. As in chapter 3, sections are included on error estimates, the hypercircle and bivariational bounds. Chapter 6 gives non-linear applications, most of which have already been mentioned in this survey.

I.6 The influence of the development of dual extremum principles on the magnetohydrodynamic pipe flow problem

In this section we intend to show how the development of dual extremum principles, as detailed in the last four sections, has influenced the work carried out on one particular problem in magnetohydrodynamics. This problem was referred to previously in section I.5.

Consider the problem of an electrically conducting viscous liquid flowing down a long straight pipe. Assume that the flow is rectilinear. A section of the pipe is shown in figure (I.6.1).

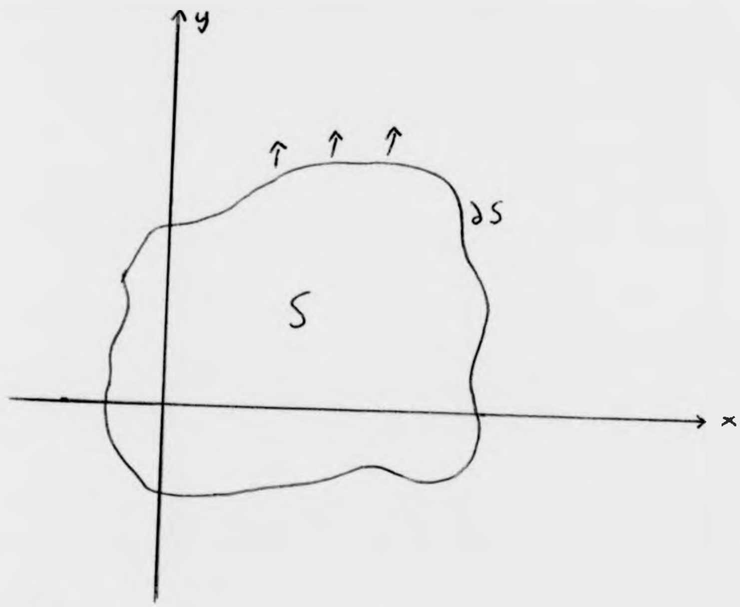


Figure I.6.1

The fluid occupies the region S which is bounded by ∂S, and the flow is pressure-driven and perpendicular to the (x,y) plane. The fluid adheres to the boundary ∂S. Suppose that a uniform magnetic field is applied in the y direction. The motion of the fluid induces a magnetic field in the direction of the flow and the Lorentz force affects the fluid velocity. If w is the fluid velocity and b is the induced field, both in suitable dimensionless form, then w and b are determined by the coupled equations

$$\nabla^2 w + M\frac{\partial b}{\partial y} = -1 \text{ in } S \tag{I.6.1}$$

$$\nabla^2 b + M\frac{\partial w}{\partial y} = 0 \text{ in } S \tag{I.6.2}$$

$$w = b = 0 \text{ on } \partial S \tag{I.6.3}$$

The first paper we shall look at is one by Smith from 1971, (62). In this paper bounds are found for the quantity

$$Q = \int_S w \, ds \tag{I.6.4}$$

Q is essentially the mass flow rate down the pipe.

The method used involves finding identities between approximate and exact solutions of the problem specified by equations (I.6.1) to (I.6.3), and then finding bounds to Q by using the Cauchy-Schwarz inequality, Greens Theorem in the plane and similar results.

The bounds obtained in this paper are

(i) $$\left| Q - \tfrac{1}{2}\int_S \left\{ (\underline{v}^x)^2 + (\text{grad } q^x)^2 \right\} ds \right| \leqslant$$
$$\tfrac{1}{2}\left[\int_S \left\{ (\underline{v}^x + \text{grad } q^x)^2 \right\} ds \int_S \left\{ (\underline{v}^x - \text{grad } q^x)^2 \right\} ds \right]^{\frac{1}{2}} \tag{I.6.5}$$

(ii) $$\left| Q^{\frac{1}{2}} - \left\{ \int_S (\text{grad } q^x)^2 \, ds \right\}^{\frac{1}{2}} \right| \leqslant \left\{ \int_S (\underline{v}^x - \text{grad } q^x)^2 \, ds \right\}^{\frac{1}{2}}$$

and (I.6.6)

(iii) $$\left| Q^{\frac{1}{2}} - \left\{ \int_S (\underline{v}^x)^2 \, ds \right\}^{\frac{1}{2}} \right| \leqslant \left\{ \int_S (\underline{v}^x - \text{grad } q^x)^2 \, ds \right\}^{\frac{1}{2}} \tag{I.6.7}$$

where, in each case, $\underline{v}^x$ and q^x satisfy the equation

$$\text{div } \underline{v}^x + M\frac{\partial q^x}{\partial y} = -1 \text{ in } S, \quad q^x = 0 \text{ on } \partial S \tag{I.6.8}$$

In paper (63), (1972), Smith obtains more accurate bounds for Q; similar methods are used to those in (62). The bounds obtained are

(i) $$Q \geqslant \frac{\left\{ \int_S q^x \, ds \right\}^2}{\int_S \left\{ (\underline{v}^x + \text{grad } q^x)^2 \right\} ds} \tag{I.6.9}$$

where $\text{div } \underline{v}^x + M\frac{\partial q^x}{\partial y} = 0$ in S, $q^x = 0$ on ∂S (I.6.10)

and

(ii) $$Q \leq \frac{1}{2} \int_s \left\{ (\underline{v}^x)^2 + (\text{grad } q^x)^2 \right\} ds$$
$$+ \frac{1}{2} \left\{ \left[\int_s \left[(\underline{v}^x)^2 + (\text{grad } q^x)^2 \right] ds \right]^2 - 4 \left[\int_s \underline{v}^x \cdot \text{grad } q^x \, ds \right]^2 \right\}^{\frac{1}{2}} \quad \text{(I.6.11)}$$

where $\operatorname{div} \underline{v}^x = M \dfrac{\partial q^x}{\partial y} = -1$ in S, $q^x = 0$ on ∂S (I.6.12)

In these two papers, the extremum principles are effectively trial and error in that, for instance, an upper bound Q^x is found which can be proved to satisfy $Q^x - Q \geq 0$.

In paper (61), by Sloan in 1973, we see the first mention of a functional as a means to obtaining extremum principles for the magnetohydrodynamic pipe flow problem.

The minimum principle used in (61) is the classic principle which was reviewed in section I.2 of this Survey. If we have an operator equation

$$A\underline{u}_e = f \quad \text{(I.6.13)}$$

then, if A is a positive-definite symmetric operator,

$$- \langle \underline{u}_e, f \rangle \leq F_1(\underline{u}) = \langle A\underline{u} - 2f, \underline{u} \rangle \quad \text{(I.6.14)}$$

where $\underline{u}$ is any element in the domain of A.

The vectors and operator used by Sloan are

$$A = \begin{bmatrix} -\nabla^2 & -M\dfrac{\partial}{\partial y} \\ M\dfrac{\partial}{\partial y} & \nabla^2 \end{bmatrix}, \quad f = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \underline{u}_e = \begin{bmatrix} w \\ b \end{bmatrix} \text{ and } \underline{u} = \begin{bmatrix} W \\ B \end{bmatrix} \quad \text{(I.6.15)}$$

with the inner product given by the equation

$$\langle \underline{u}, \underline{v} \rangle = \int_s \underline{u}^t \underline{v} \, ds \quad \text{(I.6.16)}$$

The maximum principle given in (61) is

$$- \langle \underline{u}_e, f \rangle \geq F_2(\underline{u}) \quad \text{(I.6.17)}$$

where $F_2(\underline{u}) = \langle \underline{u}, L\,\underline{u} \rangle - 2 \langle \underline{u}, f \rangle$ (I.6.18)

and $\langle \underline{u}, L\,\underline{u} \rangle = \int_s \left[2W^x - \left\{ (\text{grad } W^x)^2 + (\text{grad } B^x)^2 \right\} \right] ds$ (I.6.19)

$$\underline{u} = \begin{bmatrix} W^x \\ B^x \end{bmatrix} \text{ must satisfy } \nabla^2 W^x + M \frac{\partial B^x}{\partial y} = 0 \text{ in } S \tag{I.6.20}$$

as well as $W^x = B^x = 0$ on ∂S. The paper uses Green's theorem and a series expansion for $\underline{u}$ to show that $F_2(\underline{u})$ provides a maximum.

The bounds produced using these minimum and maximum principles are

$$- \int_S \left\{ (\text{grad } W^x)^2 + (\text{grad } B^x)^2 \right\} ds$$

$$- \int_S \left\{ (\text{grad } w)^2 + (\text{grad } b)^2 \right\} ds \leq$$

$$\int_S \left\{ (\text{grad } W)^2 + (\text{grad } B)^2 - 2W \right\} ds$$

where $\begin{bmatrix} W^x \\ B^x \end{bmatrix}$ satisfies $W^x = B^x = 0$ on ∂S and equation (I.6.20), and $\begin{bmatrix} W \\ B \end{bmatrix}$ satisfies $W = B = 0$ on ∂S.

In paper (64) (1976), Smith applies the methods from Noble and Sewell's 1972 paper ((50)) and Sewell's 1973 paper ((58)) to find saddle functionals from which dual extremum principles for the magnetohydrodynamic pipe flow problem can be found (these two papers were reviewed in sections I.4 and I.5).

Using the method from (64), in which the problem given by equations (I.6.1) to (I.6.3) is represented by the equations

$$T^x \phi = \frac{\partial H}{\partial \psi} \tag{I.6.21}$$

and
$$T \psi = \frac{\partial H}{\partial \phi} \tag{I.6.22}$$

Smith obtains the functionals

$$\langle \phi, T\psi \rangle = \int_{\partial S} \left\{ \underline{u} \cdot \underline{v} - w M \frac{\partial b}{\partial y} \right\} ds \tag{I.6.23}$$

and

$$H(\phi, \psi) = \int_{\partial S} \left\{ w + \tfrac{1}{2} (\underline{u}^2 - \underline{v}^2) + \underline{u} \cdot \underline{v} + w \text{ div } \underline{u} - b \text{ div } \underline{v} \right\} ds \tag{I.6.24}$$

where $\phi = (w \ \underline{v})$ and $\psi = (b \ \underline{u})$ (I.6.25)

Applying equations (I.6.21) and (I.6.22) to these last three equations produces

$$\begin{pmatrix} M\dfrac{\partial w}{\partial y} = -\operatorname{div} \underline{v} \\ \underline{u} = \operatorname{grad} w \end{pmatrix} \tag{I.6.26}$$

and

$$\begin{pmatrix} -M\dfrac{\partial b}{\partial y} = \operatorname{div} \underline{u} + 1 \\ \underline{v} = \operatorname{grad} b \end{pmatrix} \tag{I.6.27}$$

Eliminating $\underline{u}$ and $\underline{v}$ from equations (I.6.26) and (I.6.27) results in equations (I.6.1) and (I.6.2).

The method in (50) consists of finding a saddle functional $L(\phi, \psi)$ such that the two equations $\dfrac{\partial L(\phi,\psi)}{\partial \psi} = 0$ and $\dfrac{\partial L(\phi,\psi)}{\partial \phi} = 0$ give the required problem.

The saddle functional Smith uses in this paper is

$$L(\phi,\psi) = \int_s \left\{ \tfrac{1}{2}\left(\underline{u}^2 - \underline{v}^2\right) + w + M\, w\frac{\partial b}{\partial y} + w \operatorname{div} \underline{u} - b \operatorname{div} \underline{v} \right\} ds \tag{I.6.28}$$

Taking the gradients gives

$$\frac{\partial L}{\partial \psi} = \begin{pmatrix} -M\dfrac{\partial w}{\partial y} - \operatorname{div} \underline{v} \\ \underline{u} - \operatorname{grad} w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{I.6.29}$$

$$\frac{\partial L}{\partial \phi} = \begin{pmatrix} M\dfrac{\partial b}{\partial y} + \operatorname{div} \underline{u} + 1 \\ -\underline{v} + \operatorname{grad} b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{I.6.30}$$

Equations (I.6.1) and (I.6.2) are obtained if $\underline{u}$ and $\underline{v}$ are eliminated from equations (I.6.29) and (I.6.30).

Swetits and Rogers in 1978 (72) used the theory in Arthurs 1970 monograph (7) to find dual extremum principles for the magnetohydrodynamic pipe flow problem. The results obtained were

$$\left\{ \int_S \left[2w_1 - (\text{grad } b_1)^2 - (\text{grad } w_1)^2 \right] ds + 2\oint_{\partial S} \left[w_1 \frac{\partial w_1}{\partial x} dy - w_1 \frac{\partial w_1}{\partial y} \right] dx \right\}$$

$$\leq \int_S w \, ds \leq$$

$$\int_S \left\{ (\text{grad } b_2)^2 + w_2^2 \right\} ds \qquad (I.6.31)$$

where $\nabla^2 b_1 + M \frac{\partial}{\partial y} w_1 = 0$ in S, $w_1 = b_1 = 0$ on ∂S (I.6.32)

and $\text{div } w_2 + M \frac{\partial b_2}{\partial y} = -1$ in S, $w_2 = b_2 = 0$ on ∂S (I.6.33)

The last paper we are going to consider is one by Smith in 1979 (66). In this paper, Smith again uses the Noble and Sewell (paper (50)) method, as we discussed earlier. The bounds obtained were

$$\int_S \left\{ -\tfrac{1}{2} \underline{u} \cdot \underline{u} - \tfrac{1}{2} (\text{grad } b_1)^2 \right\} ds$$

$$\leq -\tfrac{1}{2} \int_S w \, ds \leq$$

$$\int_S \left\{ \tfrac{1}{2} \underline{v} \cdot \underline{v} + \tfrac{1}{2} (\text{grad } w_1)^2 - w \right\} ds \qquad (I.6.34)$$

where $M \frac{\partial b_1}{\partial y} + \text{div } \underline{u} = -1$ in S, $b_1 = 0$ on ∂S (I.6.35)

and $M \frac{\partial w_1}{\partial y} + \text{div } \underline{v} = 0$ in S, $w_1 = 0$ on ∂S (I.6.36)

The major effect of the development of dual extremum principles on the magnetohydrodynamic pipe flow problem seems to be that the bounds can be found more systematically. In section III.10 of this thesis dual extremum principles for the problem will be found using a particularly simple saddle functional.

CHAPTER I

I.7 Comparison Operators

This method involves replacing a linear operator by a larger or smaller operator, as appropriate, in the complementary bounds. The method could be useful for problems containing a complicated operator which can be replaced by a simpler one.

The development of the method began with a paper by Walpole in 1974 (75). This paper is concerned with a class of linear problems described by the equation

$$Au = f \tag{I.7.1}$$

where A is a symmetric, positive-definite linear operator. Starting with the classical minimum functional theorem (see section I.2), given by the equation

$$-\langle u, f\rangle \leq \langle u_1, Au_1 - 2f\rangle \tag{I.7.2}$$

where u_1 belongs to the same class of functions as u, Walpole derived the complementary bounds

$$\langle 2f - Au_1, u_1\rangle + \langle A_o u_2, u_2\rangle \leq \langle u, f\rangle \leq \langle f, u_1\rangle + \langle f - Au_1, u_1 + u_2\rangle \tag{I.7.3}$$

where A_o is a positive-definite linear symmetric operator such that

$$A_o \geq A \tag{I.7.4}$$

and u_2 is a function belonging to the same class of functions as u_1 such that

$$A_o u_2 = f - Au_1 \tag{I.7.5}$$

The lower bound in equation (I.7.3) is generally sharper than that given by (I.7.2).

Smith in 1978 (65) examined Walpole's extremum principles in the context of dual extremum principles. He started with the convex-concave saddle functional

$$L(\psi, \phi) = \langle\psi, T\phi\rangle + \tfrac{1}{2}\langle\psi, P\psi\rangle - \tfrac{1}{2}\langle\phi, R\phi\rangle - \langle f, \psi\rangle - \langle g, \phi\rangle \tag{I.7.6}$$

where P and R are symmetric positive linear operators and the linear operator T has an adjoint T^{α}.

After obtaining the dual extremum principles pertaining to equation (I.7.6), Smith introduces two further positive, symmetric linear operators, P_o and R_o, such that

$$P \geqslant P_o \quad \text{and} \quad R \geqslant R_o \tag{I.7.7}$$

Using equations in Walpole's paper, Smith derives two upper and two lower bounds, one of each offering an improvement over the dual extremum principles. The paper ends with three examples:

(a) $\nabla^2\psi = -1$ in τ, $\psi = 0$ on σ. P is taken as $-\nabla^2$ in τ, with $P_o = -\partial^2/\partial x^2$ in τ

(b) Flow in thin tubes.

(c) Fredholms integral equation.

In a 1981 paper by Smith, (68), the method was extended to non-linear problems. After summarising the theory of dual extremum principles, it was shown that if $L_a(\phi,\psi)$ and $L_b(\phi,\psi)$ are two saddle functionals stationary at (ϕ_a,ψ_a) and (ϕ_b,ψ_b) respectively, and

$$L_b(\phi,\psi) \leq L_a(\phi,\psi) \quad \forall\,(\phi,\psi) \tag{I.7.8}$$

then $L_b(\phi_b,\psi_b) \leq L_a(\phi_a,\psi_a)$ (I.7.9)

Furthermore, if $L(\phi,\psi)$ is a saddle functional with a stationary point at (ϕ_e,ψ_e), and

$$L_b(\phi,\psi) \leq L(\phi,\psi) \leq L_a(\phi,\psi) \quad \forall\,(\phi,\psi) \tag{I.7.10}$$

then $L_b(\phi_b,\psi_b) \leq L(\phi_e,\psi_e) \leq L_a(\phi_a,\psi_a)$ (I.7.11)

After defining 'joint convexity', the paper goes on to give bounds to $L(\phi_e,\psi_e)$ which are improvements on the standard dual extremum principles provided, among other conditions, $L_a(\phi,\psi) - L(\phi,\psi)$ (or $L(\phi,\psi) - L_b(\phi,\psi)$) is jointly convex. Similar conditions are given when the term $L(\phi,\psi) - L_a(\phi,\psi)$ (or $L_b(\phi,\psi) - L(\phi,\psi)$) is jointly convex.

The bounds are applied to the functional given by equation (I.7.6), with associated functionals

$$L_a(\phi, \psi) = \langle\psi, T\phi\rangle + \tfrac{1}{2}\langle\psi, P_a\psi\rangle - \tfrac{1}{2}\langle\phi, R_a\phi\rangle - \langle f,\psi\rangle - \langle g,\phi\rangle \qquad (I.7.12)$$

$$L_b(\phi, \psi) = \langle\psi, T\phi\rangle + \tfrac{1}{2}\langle\psi, P_b\psi\rangle - \tfrac{1}{2}\langle\phi, R_b\phi\rangle - \langle f,\psi\rangle - \langle g,\phi\rangle \qquad (I.7.13)$$

Taking P_a, P_b, R_a, P_b such that P_a-P, $R-R_a$, $P-P_b$ and R_b-R are all positive operators gives the results obtained in (65).

The theory is finally applied to the non-linear problem

$$\nabla^2\psi = F'(\psi) \text{ in } \tau, \quad \psi = \psi_a \text{ on } \sigma.$$

Chapter V of this thesis is concerned with comparison functionals and extends the work carried out by Smith (68).

I.8 Bivariational Bounds

If we are given an equation

$$A\phi = f, \quad f \in \mathcal{H} \qquad \text{(I.8.1)}$$

where A is an operator which can be linear or non-linear, we can very often find upper and lower bounds to the quantity $\langle\phi, f\rangle$ by using dual extremum principles; these are alternatively called complementary variational bounds. In their 1974 paper (18), Barnsley and Robinson assumed that there exists a pair (ψ, g), $g \in \mathcal{H}$, such that $A\psi = g$ and A is a linear, self-adjoint, positive-definite operator, for which complementary variational bounds can also be found to the quantity $\langle\psi, g\rangle$. Then, by combining the two sets of complementary variational bounds, complementary bivariational bounds can be found for the quantity $\langle\phi, g\rangle$. The term 'bivariational' refers to the fact that two sets of complementary variational bounds are used to obtain the bounds on $\langle\phi, g\rangle$.

The function g can be any arbitrary function belonging to $\mathcal{H}$; however, a suitable choice for g can give pointwise bounds on ϕ. An example for which pointwise bounds can be obtained will be given later in this section.

Later papers by Barnsley and Robinson and others extended the method to problems in which the operator A was non-self-adjoint or non-linear. In this section we will review the results for four different types of problems; these are

(a) A a self-adjoint, positive-definite, linear operator.

(b) A a non-self-adjoint linear operator which is bounded below.

(c) A a non-linear operator with $f = 0$.

(d) An initial value problem.

(a) <u>A a positive-definite, self-adjoint linear operator</u>

This problem was considered in papers in 1974 (18) and 1979 (55) by Barnsley and Robinson. Cole and Pack in 1975 (32) gave bivariational bounds for the same problem which were different from those of Barnsley and Robinson; and Walpole in 1974 (75) also developed bounds for $\langle \phi, g \rangle$ by introducing a comparison operator A_0. The later results of Barnsley and Robinson from (55), which offered an improvement on those in (18) are summarised below, as an example of the structure of the bounds which can be obtained on $\langle \phi, g \rangle$.

It is assumed that the operator A is bounded above and below, that is there exist real numbers α and β such that

$$0 \leq \beta I \leq A \leq \alpha I \tag{I.8.2}$$

The complementary bivariational bounds are then

$$J(\psi_b, \phi_b) + \lambda S(\psi_b, \phi_b) - \delta C(\psi_b, \phi_b)$$
$$\leq \langle \phi, g \rangle \leq$$
$$J(\psi_a, \phi_a) + \lambda S(\psi_a, \phi_a) + \delta C(\psi_a, \phi_a) \tag{I.8.3}$$

where $J(\psi, \phi) = -\langle \psi, A\phi \rangle + \langle \psi, f \rangle + \langle g, \phi \rangle$ (I.8.4)

$$S(\psi, \phi) = \langle A\phi - f, A\psi - g \rangle \tag{I.8.5}$$

$$C(\psi, \phi) = \|A\psi - g\| \, \|A\phi - f\|$$

$$\lambda = \tfrac{1}{2}\left(\frac{1}{\beta} + \frac{1}{\alpha}\right) \tag{I.8.6}$$

and $$\delta = \tfrac{1}{2}\left(\frac{1}{\beta} - \frac{1}{\alpha}\right) \tag{I.8.7}$$

The example discussed in (18) and (55) is the Fredholm equation of the second kind:

$$\phi(x) + \lambda \int_a^b k(x, y)\, \phi(y)\, dy = f(x) \tag{I.8.9}$$

where $k(x,y)$ is a symmetric kernal. Complementary bivariational bounds can be found for the quantity

$$\langle \phi, g \rangle = \int_a^b \phi(x)\, g(x)\, dx \tag{I.8.10}$$

If g is chosen to be the kernal k(x, y) then

$$\langle \phi, g \rangle = \int_a^b \phi(x)\, k(x, y)\, dx \tag{I.8.11}$$

$$= \frac{1}{\lambda}(f(y) - \phi(y)) \tag{I.8.12}$$

from (I.8.9), thus giving pointwise bounds to $\phi(y)$. In (35), numerical results are given for a particular k(x, y).

In a 1975 paper by Barnsley, (16), the method was applied to the Schrodinger equation

$$(-\nabla^2/2 + V(r))\, u(r) = E(r), \text{ with } \|u\| = 1 \tag{I.8.13}$$

and in (20) (Barnsley and Robinson, 1976) an adopted method is applied to the two-point boundary value problem specified by

$$-\frac{d}{dx}\left\{p(x)\frac{d\phi}{dx}\right\} + q(x)\,\phi(x) = r(x), \quad a \leq x \leq b \tag{I.8.14}$$

$$\alpha_1\, \phi(a) - \alpha_2\, \phi'(a) = \alpha_3 \tag{I.8.15}$$

$$\beta_1\, \phi(b) + \beta_2\, \phi'(b) = \beta_3 \tag{I.8.16}$$

The method assumes that there exists another self-adjoint operator A_-, $D(A_-) \supseteq D(A)$, such that

$$0 < \frac{1}{\mu}\langle x, x\rangle \leq \langle x, A_- x\rangle \leq \langle x, Ax\rangle \tag{I.8.17}$$

The complementary bivariational bounds are then

$$J(\phi_b, \psi_b) - (C_f(\phi_b)\, C_g(\psi_b))^{\frac{1}{2}} + \max\{0, S(\phi_b, \psi_b)\}$$
$$\leq \langle \phi, g\rangle \leq \tag{I.8.18}$$
$$J(\phi_a, \psi_a) + (C_f(\phi_a)\, C_g(\psi_a))^{\frac{1}{2}} + \min\{0, S(\phi_a, \psi_a)\}$$

where $J(\phi, \psi) = -\langle \phi, A\psi\rangle + \langle \phi, f\rangle + \langle \psi, g\rangle$ (I.8.19)

$$C_f(\phi) = \langle A\phi - f, (A_-)^{-1}(A\phi - f)\rangle \tag{I.8.20}$$

$$C_g(\psi) = \langle A\psi - g, (A_-)^{-1}(A\psi - g)\rangle \tag{I.8.21}$$

and $S(\phi, \psi) = \langle A\phi - f, (A_-)^{-1}(A\psi - g)\rangle$ (I.8.22)

Using these bounds, and via a Green's function representation for the operator $-\frac{d}{dx}\left\{p(x)\frac{d}{dx}\right\}$, bivariational bounds on $\phi(y)$ are found. The method is illustrated in (20) by three different examples.

(b) A a non-self-adjoint bounded below, closed linear operator

Bivariational bounds for this type of operator were developed in a 1975/76 paper by Barnsley and Robinson (19); Cole in 1980 (31) also considered non-self-adjoint operators. Assuming that A satisfies

$$\langle A\phi , A\phi\rangle \geq a^2\langle \phi, \phi\rangle \tag{I.8.23}$$

the bivariational bounds are:

$$J(\phi_b, \psi_b) - \frac{1}{a} C(\phi_b, \psi_b) \leq \langle \phi, g\rangle \leq J(\phi_a, \psi_a) + \frac{1}{a} C(\phi_a, \psi_a) \tag{I.8.24}$$

where $J(\phi, \psi) = - \langle \psi, A\phi\rangle + \langle \psi, f\rangle + \langle \phi, g\rangle$ (I.8.25)

and $C(\phi, \psi) = \|A\phi - f\| \, \|A^x\psi - g\|$ (I.8.26)

The paper by Barnsley and Robinson also includes applications to integral equations, and the diffusion equation

$$\frac{\partial\phi}{\partial t}(\underline{r}, t) + (p - \sigma^2 \nabla^2) \phi(\underline{r}, t) = P(\underline{r}, t),$$

$0 \leq t \leq T, \ \underline{r} \in V,$

where $\phi(\underline{r}, t) = 0$ on ∂V, $\phi(\underline{r}, 0) = 0$

(c) F a non-linear operator, with f = 0

The problem $F\phi = 0$ (I.8.27)

where F is a non-linear operator, is discussed in a 1977 paper by Barnsley and Robinson (21). Bivariational bounds given for the problem assume that there exists a positive constant c and a non-negative constant K such that

$$\|F\phi_1 - F\phi_2\| \geq c\|\phi_1 - \phi_2\| \quad \forall \phi_1, \phi_2 \in D(F_a) \subset D(F) \tag{I.8.28}$$

and $\|F\phi_1 - F\phi_2 - F'\phi(_1)(\phi_1 - \phi_2)\| \leq \frac{1}{2} K\|\phi_1 - \phi_2\|^2$

$$\forall \phi_1, \phi_2 \in D(F_a) \tag{I.8.29}$$

where F' is the functional derivative of F; F' has an adjoint F'^x such that

$$F'(\phi)^x\psi = g \tag{I.8.30}$$

The bivariational bounds are then

$$J(\bar\Psi, \bar\Phi) - C(\bar\Psi, \bar\Phi) \leq \langle \phi, g\rangle \leq J(\bar\Psi, \bar\Phi) + C(\bar\Psi, \bar\Phi)$$

$$\forall \bar\Phi \in D(F_a) \text{ and } \forall \bar\Psi \in D(F'^x) \tag{I.8.31}$$

where $J(\bar{\Psi},\bar{\Phi}) = -\langle\bar{\Psi}, F\bar{\Phi}\rangle + \langle\bar{\Phi}, g\rangle$ (I.8.32)

and $C(\bar{\Psi},\bar{\Phi}) = \frac{1}{d} \, ||F'(\bar{\Phi})^x\bar{\Psi} - g|| \, ||F\bar{\Phi}|| + \frac{K}{2c^2} \, ||\bar{\Psi}|| \, ||F\bar{\Phi}||^2$ (I.8.33)

The paper also includes a section on existance theorems and conditions for bounds. Three examples are presented, with numerical results:

(i) Bounds on the positive solution of an algebraic equation $F\phi = \phi^4 - \phi - 1 = 0$ (see section 10 of this chapter).

(ii) An integral equation from communication theory.

(iii) Bounds on the heat contained in a bar, represented by the equations

$$-\phi''(x) + \phi + \phi^2/4 = 0, \ -1 \leq x \leq 1; \quad \phi(-1) = \phi(1) = 1$$

(d) An initial value problem

Papers by Collins in 1976 and 1977 ((33) and (34)) used the 'mirror method' for initial value problems; this involves embedding the given problem in a two-point boundary value problem. Starting with the equation

$Au = Lu + Ru = f, \quad u(a) = u_a$ (I.8.34)

where L is a linear positive-definite symmetric operator and R is a skew-symmetric linear operator such that

$R^x = - R,$ (I.8.35)

the adjoint equation

$A^x v = Lv - Rv = g, \quad v(b) = \sigma u(b) + Va$ (I.8.36)

is introduced, where $\sigma \in \mathbb{R}$ and $u(b)$ is the solution u of equation (I.8.34) at t = b.

Adding and subtracting (I.8.34) and (I.8.36) gives the pair of equations

$Lw_1 + Rw_2 = h_1$ (I.8.37)

$Lw_2 + Rw_1 = h_1$ (I.8.38)

with $w_1(a) + w_2(a) = u_a$, (I.8.39)

$$(1 - \sigma)w_1(b) - (1 + \sigma)w_2(b) = V_a \qquad \text{(I.8.40)}$$

where $w_1 = \frac{1}{2}(u + v)$, $w_2 = \frac{1}{2}(u - v)$, $h_1 = \frac{1}{2}(f + g)$ and

$$h_2 = \frac{1}{2}(f - g) \qquad \text{(I.8.41)}$$

Dual extremum principles for the system defined by equations (I.8.37) to (I.8.41) are then obtained, which provide bivariational bounds to the quantity $\langle u, g \rangle$.

Applications treated in these papers include

(i) The heat equation $\frac{\partial u}{\partial t}(x, t) = \nabla^2 u(x, t) + f(x, t)$,

with $u(x, o) = u_o(x)$ and $u(x, t) = p(x, t)$ on $\partial\Omega$, where $x \in \Omega$ and $t \in (0, T)$.

(ii) The equation of unsteady flow of a viscous fluid,

$\nabla^2 u + \frac{du}{dt} = f$.

CHAPTER I

I.9 Iterative Methods

As one of the concerns of this thesis is the application of iterative methods in dual extremum principles, a survey of the literature to find previous applications has been made. Iterative methods seem, by and large, to have been ignored; the few papers containing iterative methods are reviewed here.

A 1968 paper by Robinson and Arthurs, (54), discusses how complementary bounds for Fredholm integral equations can be improved by iterative methods. It is stated, without proof, that an iterative sequence for the equation

$$\phi(r) = f(r) + \lambda K\phi(r) \qquad \text{(I.9.1)}$$

(where $K\phi(r) = \int_V K(r, s)\phi(s)\, ds$, $K(r,s)$ a symmetric kernal (I.9.2)

and $\iint_V K^2(r, s)\, dr\, ds < \infty$) (I.9.3)

defined by

$$\phi_{n+1}(r) = f(r) + \lambda K\phi_n(r) \qquad \text{(I.9.4)}$$

converges; and there exist iterative sequences

(i) For $\lambda > 0$: $G(\phi_1) \geq J(\phi_1) \geq G(\phi_2) \geq \ldots \geq I(\phi_1, u)$ (I.9.5)

(ii) For $\lambda < 0$, $|\lambda| < \lambda_1$ where λ_1, is the smallest eigenvalue of K:

$$G(\phi_1) \leq G(\phi_2) \leq \ldots \leq I(\phi_1, u) \leq \ldots \leq J(\phi_2) \leq J(\phi_1) \qquad \text{(I.9.6)}$$

$I(\phi_1, u)$, $G(\phi)$ and $J(\phi)$ are given by

$$I(\phi, u) = \tfrac{1}{2\lambda}\int_V f\phi\, dr \qquad \text{(I.9.7)}$$

$$G(\phi) = \int_V \{-\tfrac{1}{2}\phi K\phi + \tfrac{1}{\lambda}(\tfrac{1}{2}\phi^2 - f\phi)\}\, dr \qquad \text{(I.9.8)}$$

$$J(\phi) = \int_V \{\tfrac{1}{2}\phi K\phi - \tfrac{1}{\lambda}(f + \lambda K\phi)^2\}\, dr \qquad \text{(I.9.9)}$$

Robinsons review of 1971 (53) briefly considers an iterative scheme for the equation

$$L\phi + f(\phi) = 0 \text{ in } V, \quad \phi = \alpha \text{ on } \partial V \qquad \text{(I.9.10)}$$

where $f(\phi) = F'(\phi)$, $f'(\phi) > 0 \ \forall \phi$ and f^{-1} exists.

CHAPTER I

The simple cobweb iterative scheme is

$$L\Phi_{n+1} + f(\Phi_n) = 0, \quad \Phi_{n+1} = \alpha \text{ on } \partial V \tag{I.9.11}$$

and, with complementary bounds,

$$G(T\Phi) \leq -\tfrac{1}{2}\langle \Phi, f(\Phi)\rangle_V + \langle 1, F(\Phi)\rangle_V + \tfrac{1}{2}\langle \sigma^x T\Phi, \alpha\rangle_{\partial V} \leq J(\Phi) \tag{I.9.12}$$

the limit as n tends to infinity of the difference

$$J(\Phi_{n+1}) - G(T\Phi_{n+1}) = \langle 1, F(\Phi_{n+1}) - F(\Phi_n)\rangle_V - \langle F(\Phi_n), \Phi_{n+1} - \Phi_n\rangle_V \tag{I.9.13}$$

is zero provided the iterations defined by equation (I.8.11) converges to the exact solution of (I.9.10). If L is a positive operator and f' is negative, with $\sigma^x = 0$, the bounds produce the sequence

$$J(\Phi_0) \geq G(T\Phi_1) \geq J(\Phi_1) \geq G(T\Phi_2) \geq \ldots \geq I(\Phi, u) \tag{I.9.14}$$

and convergence to the stationary point is expected provided that the largest eigenvalue of the operator $(f'(\Phi))^{-1} L$ is less than -1.

Smith, in 1979 (66) applied cobweb iteration to the M.H.D. system

$$\begin{aligned} \nabla^2 u + M\partial v/\partial y &= -1 \text{ in } D \\ \nabla^2 v + M\partial u/\partial y &= 0 \text{ in } D \\ u = v &= 0 \text{ on } \partial D \end{aligned} \tag{I.9.15}$$

with the iterative scheme

Choose (u_0, v_0) : $\nabla^2 v_0 + \frac{M\partial u_0}{\partial y} = 0$ in D, $u_0 = v_0 = 0$ on ∂D

Then, for n = 0,1,2,... $\nabla^2 u_{n+1} = -1 - \frac{M\partial v_n}{\partial y}$ in D, $u_{n+1} = 0$ on ∂D (I.9.16)

$\nabla^2 v_{n+1} = -M\frac{\partial u_{n+1}}{\partial y}$ in D, $v_{n+1} = 0$ on ∂D

If the dual extremum principles are

$$B_n \leq Q \leq A_n \tag{I.9.17}$$

then $A_n - B_n = \int_D \{\nabla(u_n - u_{n+1})\}^2 \, d\tau$ (I.9.18)

and it is shown using the Schwarz inequality (see Chapter 2) and Friedrich's inequality (see page 146 of Mikhlin, 1964 (44)) that $\lim_{n\to\infty} A_n - B_n = 0$ provided $M < \sqrt{\mu}$, where $\frac{1}{\mu} \leq a^2$, and the real number a is the width of D projected onto the x - axis. Comment is then made that the iterations could be useful even in cases where convergence cannot be proven.

Two papers by Burrows and Perks in 1981, (23) and (24), considered variational-iterative principles for the non-linear problem

$$A\phi = f(\phi) \tag{I.9.19}$$

where A is a linear operator and $f(\phi)$ is non-linear. It is assumed that A has a discrete spectrum; that is, there exist eigenvalues $\lambda_i \neq 0$ and eigenvectors w_i such that $Aw_i = \lambda_i w_i$, and the eigenvalues form a complete orthonormal set.

The dual extremum principles are given as

$$J(\Phi_1, \Delta_1) \leq -\langle \phi, f \rangle \leq J(\Phi_2, \Delta_2) \tag{I.9.20}$$

where $J(\Phi, \Delta) = \langle \Phi, A\Phi \rangle - 2\langle \Phi, f \rangle + \Delta \langle A\Phi - f, A\Phi - f \rangle, \Phi \in D(A)$ (I.9.21)

and $\Delta_1 \leq -1/\lambda_i, \quad \Delta_2 \geq -1/\lambda_i \quad \forall \lambda_i$ (I.9.22)

The related equation

$$A\psi_{n+1} = f(\phi_n), \; n = 0,1,2,\ldots \tag{I.9.23}$$

is then introduced, where ϕ_n is a variational approximation to ψ_{n+1} . Then, from (I.9.19) and (I.9.23),

$$A(\phi - \psi_{n+1}) = f(\phi) - f(\phi_n) = f'(V_n)\delta\phi_n \tag{I.9.24}$$

where $\delta\phi_n = \phi - \phi_n$ and $v_n = t\phi + (1-t)\phi_n, \; t \in \,]0, 1[$ (I.9.25)

((I.8.25) is obtained by using Taylor's theorem to expand $f(\phi_n)$ about ϕ).

Putting $\delta\psi_{n+1} = \phi - \psi_{n+1}$ gives

$$\delta\psi_{n+1} = A^{-1}(f'(V_n)\delta\phi_n) \tag{I.9.26}$$

It is shown in (24) that $\lim_{n\to\infty} ||\phi - \phi_{n+1}|| = 0$ provided there exists M and ε such that, for $0 < M + \varepsilon < 1$,

$$||\phi - \phi_{n+1}|| \leq (M + \varepsilon)\, ||(\phi - \phi_n)|| \tag{I.9.27}$$

M and ε are given by

$$M = ||f'(V_n)|| \,/\, |\lambda_0| \tag{I.9.28}$$

where λ_o is the eigenvalue of smallest modulus, and $\varepsilon > 0$ satisfies

$$||\psi_{n+1} - \phi_{n+1}|| \leq ||\phi - \phi_{n+1}|| \tag{I.9.29}$$

If M is too large it can be replaced by

$$M^1 = ||(1 - b)I + bT^{-1} f'(v_n)I|| \tag{I.9.30}$$

where b, chosen to make M^1 small, satisfies

$$A\psi_{n+1} = T\bar{\phi}_n + b(f(\bar{\phi}_n) - A(\bar{\phi}_n)) \tag{I.9.31}$$

The existance of a suitable ε is ensured as $||\psi_{n+1} - \phi_{n+1}||$ can be made as small as necessary by increasing the number of parameters in ϕ_{n+1} and varying these parameters.

In (23) it is suggested that the convergence of the sequence $\{\phi_n\}$ can be accelerated by letting

$$\phi_{n+2} = \phi_{n+1} + \hat{\phi}_{n+2} \tag{I.9.32}$$

where $\hat{\phi}_{n+2}$ is chosen to satisfy

$$\langle A\hat{\phi}_{n+2}, \phi_{n-1} + \phi_{n+1} - 2\phi_n\rangle = \langle\phi_n - \phi_{n+1}, A(\phi_n - \phi_{n+1})\rangle \tag{I.9.33}$$

In (24) a method for forming $\{\phi_n\}$ is introduced which involves letting w_{np} denote a trial function containing p parameters for each n; at the nth iteration the parameters are chosen so that $J(\Delta_1, w_{n+1p})$ is stationary where $f = f(w_{np})$; convergence is considered. Finally numerical calculations are given to illustrate the methods, for several problems.

The last paper to be considered in this section is a paper by Smith published in 1981 (67). This paper deals with the construction of minimising sequences for convex functionals. The work follows that of the author on comparison operators (see Section I.7).

The paper is concerned with the strictly convex functional $L(\phi)$ with a unique minimum at $\phi = \phi_o$ such that $L'(\phi_o) = 0$. The minimum principle states that if $\phi_1 \in E = D(L'(\phi))$, $\phi_1 \neq \phi_o$, then

$$L(\phi_1) > L(\phi_o) \tag{I.9.34}$$

A minimising sequence $\{\phi_n\} \in E$ is one which satisfies

$$\lim_{n \to \infty} L(\phi_n) = L(\phi_0); \quad (I.9.35)$$

this does not necessarily mean that

$$\lim_{n \to \infty} \|\phi_n - \phi_o\| = 0. \quad (I.9.36)$$

Minimising sequences are formed by the following methods:

(a) Let $L_a(\phi)$ be a functional such that

$L_a(\phi) - L(\phi)$ is convex (I.9.37)

and $L_a'(\psi_1) = L_a'(\phi_1) - L'(\phi_1)$ for $\phi_1, \psi_1 \in E$ (I.9.38)

then $L(\phi_1) \geqslant L(\psi_1) \geqslant L(\phi_0)$ (I.9.39)

A sequence $\{\phi_n\}$ is obtained by letting

$$\phi_n = \psi_{n-1} + t_{n-1}(\phi_{n-1} - \psi_{n-1}) \quad (I.9.40)$$

where t_{n-1} satisfies

$$\langle \phi_{n-1} - \psi_{n-1}, L'(\psi_{n-1} + t(\phi_{n-1} - \psi_{n-1}))\rangle = 0 \quad (I.9.41)$$

$L(\phi_n)$ is then a monotonic decreasing sequence.

(b) Let $L_a(\phi)$, $L(\phi)$ be as above, and let there exist $\epsilon \in \,]0, 1[$ such that, for any $\lambda_1, \lambda_2 \in E$,

$$\|L'(\lambda_1) - L'(\lambda_2) - L_a'(\lambda_1) + L_a'(\lambda_2)\|$$
$$\leqslant \epsilon \| L_a'(\lambda_1) - L_a'(\lambda_2)\| \quad (I.9.42)$$

The sequence $\{\phi_n\}$, $\phi_n \in E$, where

$$L_a'(\phi_{n+1}) = L_a'(\phi_n) - L'(\phi_n) \quad (I.9.43)$$

is proved to be a minimising sequence.

The results in the paper are applied to the quadratic functional

$$L(\phi) = \tfrac{1}{2}\langle \phi, A\phi\rangle - \langle \phi, f\rangle \quad (I.9.44)$$

$L_a(\phi)$ is given by

$$L_a(\phi) = \tfrac{1}{2}\langle \phi, A_a \phi\rangle - \langle \phi, f\rangle \quad (I.9.45)$$

where A_a is a self-adjoint linear operator such that

$$A_a \geqslant A \quad (I.9.46)$$

The sequences $\{\phi_n\}$, $\{\psi_n\}, \phi_n \neq \psi_n$ are defined by

$$A_a\psi_n - f = (A_a - A)\phi_n \qquad (I.9.47)$$

$$\phi_{n+1} = \psi_n + t_n (\phi_n - \psi_n) \qquad (I.9.48)$$

$$t_n = -\frac{\langle\phi_n - \psi_n, (A_a - A)(\phi_n - \psi_n)\rangle}{\langle\phi_n - \psi_n, A(\phi_n - \psi_n)\rangle} \qquad (I.9.49)$$

Putting $A_a = MI$ where $A \leq MI$ leads to the algorithm for steepest descent,

$$\phi_{n+1} = \phi_n - \frac{\|A\phi_n - f\|^2 (A\phi_n - f)}{\langle A(A\phi_n - f), A\phi_n - f\rangle} \qquad (I.9.50)$$

which converges to the unique solution ϕ_n of $A\phi = f$ provided $0 < mI \leq A$.

A complementary lower bound is constructed by introducing a functional $L_b(\phi)$ such that $L_b(\phi)$ and $L(\phi) - L_b(\phi)$ are both convex $\forall \phi \in E$, and there exists $\mu_1 \in E$ such that

$$L_b'(\mu_1) = L_b'(\phi_1) - L'(\phi_1) \quad \forall \phi_1 \in E \qquad (I.9.51)$$

then $L(\phi_0) \geq L_b(\mu_1) - L_b(\phi_1) + L(\phi_1) + \langle\phi_1 - \mu_1, L_b'(\mu_1)\rangle$ (I.9.52)

Although an iterative scheme could be set up, this is not tried, as there is no guarantee that a maximising sequence would be produced. The quadratic functional given in (I.8.45) is again considered, and an improvement using convexity as above (equation (I.8.48) is looked at. The paper ends with the non-linear example, $\nabla^2\phi = F'(\phi)$ in D, $\phi = 0$ on C.

I.10 A new derivation of a well-known result

It occasionally happens that new techniques re-derive well known results. This brief section illustrates one such rediscovery using the techniques of dual extremum principles.

In paper (21) (Barnsley and Robinson, 1977), bivariational bounds for non-linear problems were discussed (see section 7 of this chapter); one example treated was the bounds on the positive solution of the equation $F\phi = \phi^4 - \phi - 1 = 0$. As the inner product in $\mathbb{R}$ is multiplication of functions, and the norm $\| \|$ is equivalent to the modulus $| |$, the bivariational bounds can be written:

$$J - C \leq \phi \leq J + C \qquad \text{(I.10.1)}$$

where $J = \bar{\Phi} - \bar{\Psi} F \bar{\Phi}$ (I.10.2)

and $C = \frac{1}{c} \left| \frac{\bar{\Psi}}{F'(\bar{\Phi})} - 1 \right| \left| F\bar{\Phi} \right| + \frac{K}{2c^2} \left| \bar{\Psi} \right| \left| F\bar{\Phi} \right|^2$ (I.10.3)

where $\bar{\Psi} \in D(1/F'),\ \bar{\Phi} \in D(F_\alpha) \subseteq \,]0, \infty[$ (I.10.4)

c and K are chosen to satisfy the equations

$$|F\phi_1 - F\phi_2| \geq c\, |\phi_1 - \phi_2|,\ c > 0, \forall \phi_1, \phi_2 \in D(F_\alpha) \qquad \text{(I.10.5)}$$

which implies that F is one-one on $D(F_\alpha)$;

and $|F\phi_1 - F\phi_2 - F'(\phi_1)(\phi_1 - \phi_2)| \leq \frac{K}{2} |\phi_1 - \phi_2|^2$

$K \geq 0,\ \forall \phi_1, \phi_2 \in D(F_\alpha)$ (I.10.6)

Given that the equation $F\phi = 0$ has the solution on the one-one interval $(\bar{\Phi}^-, \bar{\Phi}^+) = D(F_\alpha)$, c and K are chosen as

$c = F'(\bar{\Phi}^-)$ and $K = F''(\bar{\Phi}^+)$; (I.10.7)

$\bar{\Phi}$ and $\bar{\Psi}$ are chosen to be

$$\bar{\Phi} = \tfrac{1}{2} (\bar{\Phi}^- + \bar{\Phi}^+) \qquad \text{(I.10.8)}$$

and $\bar{\Psi} = \dfrac{1}{F'(\bar{\Phi})}$ (I.10.9)

These choices for Φ and Ψ result in

$$J = \Phi - \frac{F\Phi}{F'(\Phi)} \qquad (I.10.10)$$

$$\text{and} \quad C = \frac{F''(\Phi+)\,(F(\Phi))^2}{2\,(F'(\Phi-))^2\,|F'(\Phi)|} \qquad (I.10.11)$$

The new $D(F_\alpha)$ then becomes $(J - C, J+C)$, and the process is repeated to give tighter and tighter bounds on ϕ. The process can be rewritten as an iterative sequence.

Find an interval $(J_0 - C_0, J_0 + C_0)$ on which $F(\phi)$ changes sign and is one-one. Then, for $n = 1,2,\ldots,$ $\phi \in (J_n - C_n, J_n + C_n)$ (I.10.12)

$$\text{where} \quad J_{n+1} = J_n - \frac{F(J_n)}{F'(J_n)} \qquad (I.10.13)$$

$$\text{and} \quad C_{n+1} = \frac{F''(J_n + C_n)\,(F(J_n))^2}{2\,(F'(J_n - C_n))^2\ F'(J_n)\ \ 2} \qquad (I.10.14)$$

Equation (I.10.10) is, of course, the Newton-Raphson iterative formula for finding the roots of an equation, starting with a value J_0 near to the root of $F(\phi) = 0$.

CHAPTER II

II.1 Introduction

In this chapter we give the general principles on which this thesis is based. The first part of the chapter introduces the basic concepts, including vector space axioms, inner product, operators, functionals and functional derivatives. We then consider convexity, saddle functionals, variational principles and the dual extremum principles on which this thesis depends. The chapter ends with sections on differential and integral operators, and convergence conditions. Most sections include one or more examples.

CHAPTER II

II.2 Vector Space Axioms

Definition II.2.1

A vector space V consists of a set of elements called vectors which, together with the operations of addition and scalar multiplication, satisfy the following axioms:

(a) Addition Axioms

(i) $\phi_1 + \phi_2 = \phi_2 + \phi_1 \quad \forall\ \phi_1, \phi_2 \in V$

(ii) $\phi_1 + (\phi_2 + \phi_3) = (\phi_1 + \phi_2) + \phi_3 \quad \forall\ \phi_1, \phi_2, \phi_3 \in V$

(iii) There exists a unique zero vector $0 \in V$ such that

$\phi + 0 = 0 + \phi = \phi \quad \forall\ \phi \in V$

(iv) For every vector $\phi \in V$ there exists a unique vector $(-\phi) \in V$ such that $\phi + (-\phi) = 0$.

(b) Scalar Multiplication Axioms

(i) For every $\phi \in V$ and $\alpha \in \mathbb{R}$ there exists a unique vector $\alpha\phi \in V$.

(ii) $\alpha(\beta\phi) = (\alpha\beta)\phi \quad \forall\ \phi \in V$ and $\forall\ \alpha, \beta \in \mathbb{R}$

(iii) $1\phi = \phi, \quad 0\phi = 0 \quad \forall\ \phi \in V$

(iv) $\alpha(\phi_1 + \phi_2) = \alpha\phi_1 + \alpha\phi_2 \quad \forall\ \phi_1, \phi_2 \in V$ and

(v) $(\alpha + \beta)\phi = \alpha\phi + \beta\phi \quad \forall\ \phi \in V$ and $\forall\ \alpha, \beta \in \mathbb{R}$

Example

The space C(a,b) of functions continuous on (a,b) is a vector space.

REFERENCE: (43) pp 50-52

II.3 Inner Product

Definition II.3.1

An inner product on a vector space V is a scalar-valued function $\langle \phi_1, \phi_2 \rangle$, where $\phi_1, \phi_2 \in V$, which possesses the following properties:

(i) $\langle \phi_1, \phi_2 \rangle = \langle \phi_2, \phi_1 \rangle \quad \forall \ \phi_1, \phi_2 \in V$

(ii) $\langle \phi_1, \phi_1 \rangle \geq 0$ with equality if and only if $\phi_1 = 0, \forall \ \phi_1 \in V$

(iii) $\langle \phi_1, \alpha\phi_2 + \beta\phi_3 \rangle = \alpha\langle \phi_1, \phi_2 \rangle + \beta\langle \phi_1, \phi_3 \rangle$

$\forall \ \phi_1, \phi_2, \phi_3 \in V$ and $\forall \alpha, \beta \in \mathbb{R}$

Example

An important example of an inner product is

$$\langle \phi_1, \phi_2 \rangle = \int_a^b \phi_1(t) \ \phi_2(t) \ dt, \quad a \leq t \leq b \qquad (II.3.1)$$

This inner product is used frequently in later applications.

Definition II.3.2

An Inner Product Space $X = \{V, \langle \ , \ \rangle\}$ consists of a real linear vector space V with an inner product defined on it.

REFERENCE: (77) p 170

Definition II.3.3

A real inner product space H is called a Hilbert space if the space is complete in the sense that any Cauchy sequence in H has a limit also in H. A Cauchy sequence is one such that $\lim_{m,n \to \infty} \langle x^n - x^m, x^n - x^m \rangle = 0$ where the sequence $\{x^n\} \in H \ \forall \ m,n \in \mathbb{R}$

REFERENCE: (77) pp 19 and 171

II.4 Normed Vector Space

Definition II.4.1

A **normed vector space**, denoted by $(V, ||\cdot||)$, consists of a vector space V and a norm $||\cdot||$ on V.

Definition II.4.2

A **norm** on V is a real valued function $||\phi||$ defined $\forall \phi \in V$, which satisfies the following axioms:

(i) $||\phi|| \geq 0$, with equality if and only if $\phi = 0$.

(ii) $||\phi_1 + \phi_2|| \leq ||\phi_1|| + ||\phi_2|| \quad \forall \phi_1, \phi_2 \in V$

(iii) $||\lambda\phi|| = |\lambda| \, ||\phi|| \quad \forall \phi \in V$ and $\forall \lambda \in \mathbb{R}$

Definition II.4.3

The natural norm of a vector in a real inner product space X is given by $||\phi|| = \langle\phi, \phi\rangle^{\frac{1}{2}}$. It is easy to show that it satisfies the axioms above.

Example

The norm of a vector whose inner product is given by equation (II.3.1) is

$$||\phi|| = \left\{\int_a^b (\phi(t))^2 \, dt\right\}^{\frac{1}{2}} \qquad \text{(II.4.1)}$$

REFERENCES: (45) pp 135-137; (42) pp 261-262

Lemma II.4.1

Cauchy-Schwarz Inequality: For any ϕ_1, ϕ_2 in an inner product space X,

$$|\langle\phi_1, \phi_2\rangle|^2 \leq \langle\phi_1, \phi_1\rangle \langle\phi_2, \phi_2\rangle = ||\phi_1||^2 \; ||\phi_2||^2 \qquad \text{(II.4.2)}$$

REFERENCE: (77) p 171

Lemma II.4.2

If u(x) is a continuous function with a continuous derivative for $x \in [a,b]$ and u(a) = 0, then

$$\int_a^b (u(x))^2 \, dx \leq \tfrac{1}{2} (b - a)^2 \int_a^b (u'(x))^2 \, dx \qquad \text{(II.4.3)}$$

Proof

$$\int_a^x (u'(t))\,dt = (u(t))_a^x = u(x) - u(a) = u(x) \text{ as } u(a) = 0$$

Hence $u(x) = \int_a^x (u'(t))\,dt$

Therefore

$$(u(x))^2 = \left[\int_a^x (u'(t)dt)\right]^2$$

$$\leq \left(\int_a^x (u'(t)^2\,dt)\right)\left(\int_a^x 1\,dt\right)$$

by Lemma (II.3.1) with $\phi_1 = u'(t)$ and $\phi_2 = 1$

As $(u'(t))^2$ is positive for all t,

$$\int_a^x (u'(t))^2\,dt \leq \int_a^b (u'(t))^2\,dt$$

$$\equiv \int_a^b (u'(x))^2\,dx$$

Then

$$(u(x))^2 \leq \left(\int_a^b (u'(x))^2\,dx\right)(x - a)$$

Integrate both sides of the equation on $[a,b]$:

$$\int_a^b (u(x))^2\,dx \leq \left(\int_a^b (x - a)\,dx\right)\left(\int_a^b (u'(x))^2\,dx\right)$$

$$= \tfrac{1}{2}(b - a)^2 \int_a^b (u'(x))^2\,dx$$

as required.

This is the one-dimensional version of Friedrich's inequality which is given for functions of two variables on pages 36 and 145-146 of (44).

II.5 Operators

Definition II.5.1

An operator T is a transformation from a vector space X to another vector space, say Y. We therefore write

$$T : X \rightarrow Y \tag{II.5.1}$$

or, if $\phi \in X$ and $\psi \in Y$,

$$T\phi = \psi \tag{II.5.2}$$

Definition II.5.2

An operator $T : X \rightarrow Y$ is said to be a linear operator if

$$T(\alpha\phi_1 + \beta\phi_2) = \alpha T\phi_1 + \beta T\phi_2 \tag{II.5.3}$$

$\forall \phi_1, \phi_2 \in X$ and $\forall \alpha, \beta \in \mathbb{R}$

Example

The differentiation operator $\frac{d}{dx}$ is an example of a linear operator.

Definition II.5.3

A linear operator $T : X \rightarrow Y$ has an adjoint if there exists a second operator $T^x : Y \rightarrow X$ such that

$$\langle\psi, T\phi\rangle = \langle T^x\psi, \phi\rangle + (S(\phi,\psi)) \tag{II.5.4}$$

$\forall \phi, \psi$ in the domains of T and T^x respectively, where X and Y are now inner product spaces.

The conjunct $(S(\phi,\psi))$, where it exists, generally denotes boundary terms. The term formally-adjoint is often used if boundary terms appear.

Example

An example is furnished by the differentiation operator $\frac{d}{dt}$ and the inner product given by equation (II.3.1):

Let $T = \frac{d}{dt}$; then we have, using the integration by parts formula,

$$\int_a^b \psi(t) \frac{d\phi(t)}{dt} dt = -\int_a^b \phi(t) \frac{d\psi(t)}{dt} dt + (\phi(t)\psi(t))_a^b$$

giving $T^x = -\frac{d}{dt}$, $(S(\phi,\psi)) = (\phi(t)\ \psi(t))_a^b$

REFERENCES: (9) p 18; (50) p 189

NOTE: The boundary term $(S(\phi,\psi))$ is sometimes included as part of the operator in the inner product; for instance in (65), section 5, the operator P is defined by $P = \begin{cases} -\nabla^2 & \text{in } \tau \\ \underline{n}.\nabla & \text{on } \sigma \end{cases}$

where the vector spaces are given by

$$\Phi = \begin{pmatrix} \phi \text{ in } \tau \\ \phi \text{ on } \sigma \end{pmatrix} \quad \text{and} \quad \Psi = \begin{pmatrix} \psi \text{ in } \tau \\ \psi \text{ on } \sigma \end{pmatrix}$$

and ϕ and ψ in τ are twice differentiable functions which are continuous with ϕ and ψ on σ respectively. With the inner product defined by

$$\langle \Phi_1, \Phi_2 \rangle = \int \phi_1\ \phi_2\ d\tau + \int \phi_1\ \phi_2\ d\sigma$$

we have

$$\langle \Phi, P\Psi \rangle = \int -\phi\nabla^2\psi\ d\tau + \int_\sigma \phi\underline{n}.\nabla\psi\ d\sigma$$

Definition II.5.4

The linear operator $T : X \to Y$ is said to be bounded if there exists a real positive number k such that

$$\|T\phi\| \leq k\|\phi\| \quad \forall\ \phi \in X \tag{II.5.5}$$

Example

The linear operator defined by

$$T\phi(t) = \int_0^1 K(t-\tau)\ \phi(\tau)\ d\tau, \quad t \in [0,1]$$

where K belongs to the normed space of integrable functions on (0,1), is an example of a bounded linear operator.

REFERENCE: (45) pp 144-146

Definition II.5.5

The linear operator $T : X \to Y$ is symmetric if

$$\langle T\phi_1, \phi_2 \rangle = \langle \phi_1, T\phi_2 \rangle \quad \forall\ \phi_1, \phi_2 \in X \tag{II.5.6}$$

In other words, $T = T^x$ and $X = Y$

Example 1

The differential operator $T = -\frac{d^2}{dt^2}$, with $\phi(t)$ belonging to the space of twice-differentiable functions on $(0,1)$ such that $\phi(0) = \phi(1) = 0$, is symmetric since

$$\begin{aligned}
\langle T\phi_1, \phi_2\rangle &= -\int_0^1 \frac{d^2\phi_1}{dt^2}\,\phi_2\,dt \\
&= -\int_0^1 \frac{d^2\phi_2}{dt^2}\,\phi_1\,dt + \left[\phi_1 \frac{d\phi_2}{dt} - \phi_2 \frac{d\phi_1}{dt}\right]_0^1 \\
&= -\int_0^1 \frac{d^2\phi_2}{dt^2}\,\phi_1\,dt \quad \text{as } \phi_1(0) = \phi_2(0) = \phi_1(1) = \phi_2(1) = 0 \\
&= \langle \phi_1, T\phi_2\rangle
\end{aligned}$$

REFERENCE: (45) pp 222-223

Example 2

$T^x T$ and TT^x are both symmetric operators on common domains since

$$\begin{aligned}
\langle T^x T\phi_1, \phi_2\rangle &= \langle T\phi_1, T\phi_2\rangle \\
&= \langle \phi_1, T^x T\phi_2\rangle
\end{aligned}$$

and

$$\begin{aligned}
\langle TT^x \phi_1, \phi_2\rangle &= \langle T^x\phi_1, T^x\phi_2\rangle \\
&= \langle \phi_1, TT^x\phi_2\rangle
\end{aligned}$$

Definition II.5.6

The bounded linear operator $T : X \to X$ is self-adjoint if

$$\langle T\phi_1, \phi_2\rangle = \langle \phi_1, T\phi_2\rangle \quad \forall \ \phi_1, \phi_2 \in X \tag{II.5.6}$$

Example

The integral operator $T : \mathcal{L}_2(0,1) \to \mathcal{L}_2(0,1)$

$$T\phi(t) = \int_0^1 K(t,\tau)\,\phi(\tau)\,d\tau,$$

where $\mathcal{L}_2$ (0,1) is the normed space of square integrable functions $f(x)$ on (0,1) such that $\int_0^1 (f(x))^2\,dx < \infty$ is self-adjoint if $K(t,\tau) = K(\tau,t)$

REFERENCE: (45) p 222

NOTE

In the literature, the term 'self-adjoint' is often applied to an operator which satisfies equation (II.5.6), even though it may not be bounded; for instance in (65) it is inferred that the Laplacian operator, defined by $\nabla^2\phi(x,y) = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2}$, where $\phi = 0$ on the boundary is self-adjoint. On pages 223-4 of (45) this operator is shown to be symmetric as it satisfies equation (II.5.6); but, strictly speaking it is not self-adjoint as it is unbounded. On the other hand, matrix operators are bounded but are usually called symmetric if they satisfy equation (II.5.6).

In this and the following chapters, the use of the terms 'symmetric' and 'self-adjoint' will generally refer to unbounded and bounded operators respectively.

Definition II.5.7

An operator $T : X \rightarrow X$ which satisfies equation (II.5.6) is said to be positive if

$$\langle T\phi, \phi \rangle \geq 0 \quad \forall\ \phi \in X \tag{II.5.7}$$

and positive-definite if

$$\langle T\phi, \phi \rangle \geq 0 \quad \forall\ \phi \in X \text{ with equality if and only if } \phi = 0 \tag{II.5.8}$$

'Positive' is sometimes called 'Positive Semi-Definite' in the literature.

Example 1

Example 1 following Definition (II.5.5),

$T\phi = -\frac{d^2\phi}{dt^2}$, $\phi(0) = \phi(1) = 0$, is positive-definite since

$$\langle T\phi,\phi\rangle = -\int_0^1 \frac{d^2\phi}{dt^2}.\phi \quad dt$$

$$= -\left(\phi\frac{d}{dt}\phi\right)_0^1 + \int_0^1 \left(\frac{d\phi}{dt}\right)^2 dt$$

$$= \int_0^1 \left(\frac{d\phi}{dt}\right)^2 dt$$

which is positive for all ϕ and vanishes only for $\phi(t) = 0$.

REFERENCES: (45) p 225; (77) p 214; (50) p 190

Example 2

The operators $T^x T$ and TT^x are both positive since

$$\langle\phi, T^x T\phi\rangle = \langle T\phi, T\phi\rangle$$

$$\langle\phi, TT^x\phi\rangle = \langle T^x\phi, T^x\phi\rangle$$

and, by definition (II.2.1),

$$\left.\begin{matrix}\langle T\phi, T\phi\rangle \\ \langle T^x\phi, T^x\phi\rangle\end{matrix}\right\} \geq 0 \text{ with equality if and only if } \left.\begin{matrix}T\phi \\ T^x\phi\end{matrix}\right\} = 0$$

Definition II.5.8

By analogy, an operator $T : X \to X$ which satisfies equation (II.5.6) is said to be negative (or negative semi-definite) if

$$\langle T\phi,\phi\rangle \leq 0 \quad \forall \phi \in X \qquad \text{(II.5.9)}$$

and negative-definite if

$$\langle T\phi,\phi\rangle \leq 0 \quad \forall \phi \in X \text{ with equality if and only if } \phi = 0 \qquad \text{(II.5.10)}$$

Definition II.5.9

Let A and B be symmetric or self-adjoint operators on an inner product space X; then

(a) $A \geq B$ if $\langle\phi, A\phi\rangle \geq \langle\phi, B\phi\rangle \quad \forall \phi \in X$ (II.5.11)

(b) $A > B$ if $\langle\phi, A\phi\rangle > \langle\phi, B\phi\rangle \quad \forall \phi \in X, \phi \neq 0$ (II.5.12)

REFERENCE: (77) p 215

Example

An example in which one operator is greater than another is given in (65):

Let $P\psi = -\nabla^2\psi$ in the square $\tau = \{ |x| \leq 1 , |y| \leq 1 \}$,

with $\psi = 0$ on the boundary σ;

P_0 is chosen such that $\langle \psi, P_0 \psi \rangle = \int_{-1}^{1} \int_{-1}^{1} \left(\frac{\partial \psi}{\partial x}\right)^2 dx\, dy$

$$\langle \psi, P\psi \rangle = \int_{-1}^{1}\int_{-1}^{1} -\psi \nabla^2 \psi \; dx\, dy$$

$$= \int_{-1}^{1}\int_{-1}^{1} \left\{ \left(\frac{\partial \psi}{\partial x}\right)^2 + \left(\frac{\partial \psi}{\partial y}\right)^2 \right\} dx\, dy$$

Therefore, $\langle \psi, P\psi \rangle - \langle \psi, P_0 \psi \rangle = \int_{-1}^{1}\int_{-1}^{1} \left(\frac{\partial \psi}{\partial y}\right)^2 dx\, dy$

which is greater than zero for all $\psi \neq 0$.

Hence $P > P_0$.

Definition II.5.10

Let $T : X \to Y$ be a bounded operator; then, by definition (II.5.4)

$\|T\phi\| \leq k \|\phi\| \quad \forall \phi \in X$ where k is a real number. The smallest value of k, $\|T\|$, is called the operator norm and is defined as

$$\|T\| = \sup_{\phi \neq 0} \frac{\|T\phi\|}{\|\phi\|} \qquad \text{(II.5.13)}$$

It then follows that $\|T\phi\| \leq \|T\| \|\phi\|$ (II.5.14)

REFERENCE: (45) p 147

Lemma II.5.1

If $T : X \to X$ is a self-adjoint operator, we can always find two numbers m_T and M_T such that

$$m_T \langle \phi, \phi \rangle \leq \langle \phi, T\phi \rangle \leq M_T \langle \phi, \phi \rangle \qquad \text{(II.5.15)}$$

Proof

As T is self-adjoint, it is bounded, and therefore

$$\|T\|^2 = \langle T\phi, T\phi \rangle \leq k^2 \langle \phi, \phi \rangle \qquad \text{(II.5.16)}$$

(by definition (II.5.4)).

Now, by Lemma (II.4.1),

$$|\langle \phi, T\phi \rangle| \leq \langle \phi, \phi \rangle^{\frac{1}{2}} \langle T\phi, T\phi \rangle^{\frac{1}{2}}$$

$$\leq k \langle \phi, \phi \rangle \quad \text{using equation (II.5.16)}$$

Hence $-k\langle \phi, \phi \rangle \leq \langle \phi, T\phi \rangle \leq k\langle \phi, \phi \rangle$

and equation (II.5.15) follows by letting $m_T = -k$ and $M_T = k$.

Equation (II.5.15) can alternatively be written

$$m_T I \leq T \leq M_T I \qquad \text{(II.5.17)}$$

REFERENCE: (77) p 213

<u>Theorem II.5.1</u>

If $T : X \to X$ is a self-adjoint operator, then

$$\|T\| = m = \max(|m_T|, |M_T|) \qquad \text{(II.5.18)}$$

<u>Proof</u>

Clearly, $m = \sup_{\|\phi\| = 1} |\langle T\phi, \phi \rangle| \leq \|T\|$ (II.5.19)

For any $\phi, \psi \in X$,

$$\langle T(\phi \pm \psi), \phi \pm \psi \rangle = \langle T\phi, \phi \rangle + \langle T\psi, \psi \rangle \pm \langle T\psi, \phi \rangle \pm \langle T\phi, \psi \rangle$$

Therefore,

$$\langle T(\phi + \psi), \phi + \psi \rangle - \langle T(\phi - \psi), \phi - \psi \rangle$$

$$= 2\langle T\psi, \phi \rangle + 2\langle T\phi, \psi \rangle$$

$$= 4\langle \phi, T\psi \rangle \quad \text{as } T \text{ is self-adjoint} \qquad \text{(II.5.20)}$$

Now, as $m = \sup_{\|\phi\| = 1} |\langle T\phi, \phi \rangle|$

$$|T(\phi + \psi), (\phi + \psi)| = \frac{1}{\|\phi + \psi\|^2} |\langle T(\phi + \psi), \phi + \psi \rangle| \, \|\phi + \psi\|^2 \quad (\phi \neq -\psi)$$

$$= \left|\left\langle T\left(\frac{\phi + \psi}{\|\phi + \psi\|}\right), \left(\frac{\phi + \psi}{\|\phi + \psi\|}\right)\right\rangle\right| \|\phi + \psi\|^2$$

$$\leq m \|\phi + \psi\|^2$$

and similarly,

$$|\langle T(\phi - \psi), \phi - \psi \rangle| \leq m\|\phi - \psi\|^2$$

Hence

$$4\langle \phi, T\psi \rangle = \langle T(\phi+\psi), \phi+\psi \rangle - \langle T(\phi-\psi), \phi-\psi \rangle$$
$$\leq \left| \langle T(\phi+\psi), \phi+\psi \rangle - \langle T(\phi-\psi), \phi-\psi \rangle \right|$$
$$\leq \left| \langle T(\phi+\psi), \phi+\psi \rangle \right| + \left| \langle T(\phi-\psi), \phi-\psi \rangle \right|$$
$$\leq m\|\phi+\psi\|^2 + m\|\phi-\psi\|^2$$
$$= 2m\left(\|\phi\|^2 + \|\psi\|^2\right) \qquad \text{(II.5.21)}$$

We now let $\phi = \lambda z$ and $\psi = Tz/\lambda$, for $z \in X$, $\|z\| = 1$, and λ a non-zero real number. Then

$$\langle \phi, T\psi \rangle = \langle \lambda z, TT\tfrac{z}{\lambda} \rangle = \|Tz\|^2$$

Therefore, $4\langle \phi, T\psi \rangle$

$$= 4\|Tz\|^2 \leq 2m\left(\lambda^2 \|z\|^2 + \frac{\|Tz\|^2}{\lambda^2}\right)$$

Set $\lambda^2 = \|Tz\|$, then

$$4\|Tz\|^2 \leq 2m\left(\|Tz\|\,\|z\|^2 + \frac{\|Tz\|^2}{\|Tz\|}\right)$$
$$= 2m\left(2\|Tz\|\right) \qquad \|z\| = 1$$

or $4\|Tz\| \leq 2m\,(2)$

Hence $\|Tz\| \leq m$

$$\|Tz\| \leq \|T\|\,\|z\| = \|T\| \leq m$$

so $\|T\| = m = \max\left(|m_T|, |M_T|\right)$

REF: (77) p 213

Definition II.5.11

Let T be a linear mapping from a normed vector space V to a normed linear vector space W. T is a closed operator if, for every sequence $x_n \in V$, $n = 1,2,3,\ldots$ such that $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} Tx_n = y$, then $x \in V$ and $Tx = y$.

REFERENCE: (77) p 203

Example

Let $T = \dfrac{d^m}{dt^m}$ with $D(T) = C^m(0,1)$, the space of m-times continuously differentiable functions on $(0,1)$.

CHAPTER II

Let $\{x_n\} \in D(T)$ and $\lim_{n \to \infty} x_n = x$

Now $\lim_{n \to \infty} Tx_n = \lim_{n \to \infty} \frac{d^m(x_n)}{dt^m} = y$ implies that $y \in C^m(0,1)$,

hence $\frac{d^m}{dt^m}(x_n)$ converges uniformly to y and y is continuous.

If $\frac{d^m}{dt^m}(x_n)$ converges uniformly then x is continuously differentiable and

$\frac{d^m x}{dt^m} = y$.

Hence $\frac{d^m}{dt^m}$ is a closed operator.

A similar argument is obtained if $D(T) = L_p(0,1)$, the normed space of integrable functions on (0,1).

REFERENCE: (45) p 216

Definition II.5.12

The set A is said to be dense in the set B if $A \subset B \subset \overline{A}$, where $\overline{A}$ is the closure of the set A. By 'the closure $\overline{A}$ of the set A' is meant the set of points $x \in A$ for which there exist Cauchy sequences $\{x_n\} \in A$ with

$$\lim_{n \to \infty} \langle x^n, x \rangle = 0$$

REFERENCE: (77) pp 75-76

II.6 Functionals

Definition II.6.1

A transformation L on a vector space V into the space of real numbers is said to be a functional in V. If L is a linear transformation then L is a linear functional.

Example

An example of a linear functional is given by

$$L : V \to \mathbb{R} \quad \text{where } L(\phi) = \int_0^1 \phi(t)\, dt$$

and V is the space of continuous functions $\phi(t)$ defined in (0,1).

REFERENCE: (45) pp 49-50

In the following chapters we will be concerned with functionals defined on two inner product spaces; the usual notation is given by $L : X \times Y \to \mathbb{R}$ where $X \times Y$ denotes the Cartesian product space of two inner product spaces X and Y, which need not be the same. A typical example of such a functional is given in (65):

$$L(\phi, \psi) : X \times Y \to \mathbb{R}$$

where

$$L(\phi, \psi) = \langle \psi, T\phi \rangle + \tfrac{1}{2}\langle \psi, P\psi \rangle - \tfrac{1}{2}\langle \phi, R\phi \rangle - \langle f, \psi \rangle - \langle g, \phi \rangle$$

where P and R are symmetric, positive, linear operators and T is a linear operator with an adjoint T^x.

CHAPTER II

II.7 Functional Derivatives

There are a number of differentiation operators in functional analysis. In this thesis we shall be concerned exclusively with the gradient operator in inner product spaces.

Definition II.7.1

Let X be an inner product space, and let $L(\phi)$ be a functional such that

$L(\phi) : D(L) \subseteq X \to \mathbb{R}$

Then $L(\phi)$ is said to be differentiable at ϕ if there exists an element $g \in X$ such that

$$L(\phi + h) = L(\phi) + \langle h, g \rangle + o(\| h \|) \quad \text{(II.7.1)}$$

where $h \in X$. Such an element g is called the Gateaux derivative or gradient of $L(\phi)$ and we write

$$g = \nabla L(\phi). \quad \text{(II.7.2)}$$

We then have the mapping

$\nabla L(\phi) : D(L) \subseteq X \to X$

A convenient method for finding the explicit form of the derivative is to use:

$$\left[\frac{d}{dt} L(\phi + th) \right]_{t=0} = \langle h, \nabla L(\phi) \rangle \quad \text{(II.7.3)}$$

REFERENCE: (50) p 190

Example 1

Let $L(\phi) = \langle \phi, A\phi \rangle$, where $A : X \to X$ has an adjoint A^x such that

$\langle \phi, A\psi \rangle = \langle \psi, A^x \phi \rangle$

$$\left[\frac{d}{dt} L(\phi + th) \right]_{t=0} = \left[\frac{d}{dt} \langle \phi + th, A(\phi + th) \rangle \right]_{t=0}$$

$$= \left[\frac{d}{dt} \{ \langle \phi, A\phi \rangle + \langle A\phi, th \rangle + \langle \phi, A\,th \rangle + \langle th, A\,th \rangle \} \right]_{t=0}$$

$$= \left[\langle A\phi, h \rangle + \langle \phi, A\,h \rangle + 2t \langle h, A\,h \rangle \right]_{t=0}$$

$$= \langle A\phi, h \rangle + \langle \phi, A\,h \rangle$$

$$= \langle h, (A + A^x)\phi \rangle$$

Hence $\nabla L(\phi) = (A + A^x)\phi$

CHAPTER II

REFERENCE: (9) p 23

Example 2

Let $L(\phi) : D(L) \subseteq X \to \mathbb{R}$ be defined by the equation

$$L(\phi) = \int_V F(\phi)\, dV$$

where $F(\phi)$ is a real continuous differentiable function of ϕ.

Then $\left[\frac{d}{dt} L(\phi + th)\right]_{t=0} = \left[\frac{d}{dt}\int_V F(\phi + th)\, dV\right]_{t=0}$

$$= \left[\int_V \frac{d}{dt} F(\phi + th)\, dV\right]_{t=0}$$

$$= \left[\int_V F'(\phi + th)\, h\, dV\right]_{t=0}$$

$$= \int_V F'(\phi)\, h\, dV$$

$= \langle h, \nabla L(\phi)\rangle$ using equation (II.7.3)

Hence $\nabla L(\phi) = F'(\phi)$.

Definition II.7.2

Let $L(\phi,\psi) : D(L) \subseteq X \times Y \to \mathbb{R}$ be a functional where $X \times Y$ is the product space of two inner product spaces.

By analogy with definition (II.7.1),

(a) $L(\phi,\psi)$ is differentiable at ϕ if there exists an element $g \in X$ such that

$$L(\phi + h, \psi) = L(\phi,\psi) + \langle h, g\rangle + o(\|h\|) \qquad \text{(II.7.4)}$$

where $h \in X$. The element g is called the Gateaux derivative with respect to ϕ of $L(\phi,\psi)$ and we write $g = \nabla_\phi L(\phi,\psi)$; (II.7.5)

the mapping is then $\nabla_\phi L(\phi,\psi) : D(L) \subseteq X \times Y \to X$.

(b) $L(\phi,\psi)$ is differentiable at ψ if there exists an element $k \in Y$ such that

$$L(\phi, \psi + \ell) = L(\phi,\psi) + \langle k, \ell\rangle + o(\|\ell\|) \qquad \text{(II.7.6)}$$

where $\ell \in Y$. The element k is called the Gateaux derivative with respect to ψ of $L(\phi,\psi)$ and we write $k = \nabla_\psi L(\phi,\psi)$; (II.7.7) the mapping is then $\nabla_\psi L(\phi,\psi) : D(L) \subseteq X \times Y \to Y$.

We find $\nabla_\phi L(\phi,\psi)$ and $\nabla_\psi L(\phi,\psi)$ by using

$$\left[\frac{d}{dt} L(\phi + th, \psi)\right]_{t=0} = \langle h, \nabla_\phi L(\phi,\psi)\rangle \tag{II.7.8}$$

and

$$\left[\frac{d}{dt} L(\phi, \psi + tk)\right]_{t=0} = \langle k, \nabla_\psi L(\phi,\psi)\rangle \tag{II.7.9}$$

REFERENCE: (9) pp 23-24

<u>Example 1</u>

Let $L(\phi,\psi) = \int_0^1 (\phi(x))^2 (\psi(x))^2 \, dx$, $0 \leq x \leq 1$, where the inner product is given by $\langle u,v\rangle = \int_0^1 u(x)\, v(x)\, dx$.

$$\begin{aligned}\left[\frac{d}{dt} L(\phi + th, \psi)\right]_{t=0} &= \left[\frac{d}{dt}\int_0^1 (\phi(x) + th)^2 (\psi(x))^2 \, dx\right]_{t=0} \\ &= \left[\frac{d}{dt}\int_0^1 (\phi(x))^2 + 2th\phi(x) + t^2h^2)(\psi(x))^2 \, dx\right]_{t=0} \\ &= \left[\int_0^1 (2h\phi(x) + 2th^2)(\psi(x))^2 \, dx\right]_{t=0} \\ &= \int_0^1 2h\phi(x)\,(\psi(x))^2 \, dx\end{aligned}$$

Hence $\nabla_\phi L(\phi,\psi) = 2\phi(x)\,(\psi(x))^2$

Similarly, $\nabla_\psi L(\phi,\psi) = 2\psi(x)\,(\phi(x))^2$

<u>Example 2</u>

Let $L(\phi,\psi)$ be a functional such that

$L(\phi,\psi) : D(L) \subseteq X \times Y \to \mathbb{R}$, where $X \times Y$ is a product space of two real inner product spaces.

Let

$$L(\phi,\psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B\phi\rangle - \tfrac{1}{2}\langle\psi, C\psi\rangle + \langle\phi, f\rangle + \langle\psi, g\rangle \tag{II.7.10}$$

where A, B and C are linear operators, A has an adjoint A^x and B and C are symmetric.

$$\left[\frac{d}{dt} L(\phi + th, \psi)\right]_{t=0} = \left[\frac{d}{dt}\left\{\langle \phi + th, A\psi\rangle + \tfrac{1}{2}\langle \phi + th, B(\phi + th)\rangle - \tfrac{1}{2}\langle \psi, C\psi\rangle + \langle \phi + th, f\rangle + \langle \psi, g\rangle\right\}\right]_{t=0}$$

$$= \left[\langle h, A\psi\rangle + \tfrac{1}{2}\langle \phi, Bh\rangle + \tfrac{1}{2}\langle h, B\phi\rangle + t\langle h, h\rangle + \langle h, f\rangle\right]_{t=0}$$

$= \langle h, A\psi + B\phi + f\rangle$ as B is symmetric.

Hence $\nabla_\phi L(\phi,\psi) = A\psi + B\phi + f$ (II.7.11)

$$\left[\frac{d}{dt} L(\phi, \psi + tk)\right]_{t=0} = \left[\frac{d}{dt}\left\{\langle \phi, A(\psi + tk)\rangle + \tfrac{1}{2}\langle \phi, B\phi\rangle - \tfrac{1}{2}\langle \psi + tk, C(\psi + tk)\rangle + \langle \phi, f\rangle + \langle \psi + tk, g\rangle\right\}\right]_{t=0}$$

$$= \left[\langle \phi, Ak\rangle - \tfrac{1}{2}\langle \psi, Ck\rangle - \tfrac{1}{2}\langle k, C\psi\rangle - t\langle k, k\rangle + \langle k, g\rangle\right]_{t=0}$$

$= \langle k, A^x\phi - C\psi + g\rangle$ as C is symmetric and A has an adjoint A^x

Hence $\nabla_\psi L(\phi,\psi) = A^x\phi - C\psi + g$ (II.7.12)

This functional will be used extensively in later chapters.

<u>Example 3</u>

Let X be the space $\Phi = \begin{bmatrix}\phi(x)\\ \phi(a)\end{bmatrix}$ with inner product

$$\langle \Phi_1, \Phi_2\rangle_x = \int_a^b \phi_1(x)\,\phi_2(x)\,dx + \phi_1(a)\,\phi_2(a) \tag{II.7.13}$$

$$\text{Let } L(\phi,\psi) = \int_a^b \phi(x)\,\frac{d\psi(x)}{dx}\,dx + \phi(a)\,\psi(a) \tag{II.7.14}$$

where $\phi(x)$ and $\psi(x)$ are functions which are differentiable and integrable for $x \in (a,b)$

$$\left[\frac{d}{dt} L(\phi + th, \psi)\right]_{t=0} = \left[\frac{d}{dt}\left\{\int_a^b (\phi(x) + th)\,\frac{d\psi(x)}{dx}\,dx + (\phi(a) + th)\psi(a)\right\}\right]_{t=0}$$

$$= \left[\int_a^b h\,\frac{d\psi(x)}{dx}\,dx + h\psi(a)\right]_{t=0}$$

$$\left[\frac{d}{dt} L(\phi + th, \psi)\right]_{t=0} = \left[\frac{d}{dt}\left\{<\phi + th, A\psi> + \tfrac{1}{2}<\phi + th, B(\phi + th)>\right.\right.$$
$$\left.\left. - \tfrac{1}{2}<\psi, C\psi> + <\phi + th, f> + <\psi, g>\right\}\right]_{t=0}$$
$$= \left[<h, A\psi> + \tfrac{1}{2}<\phi, Bh> + \tfrac{1}{2}<h, B\phi> + t<h, h> + <h, f>\right]_{t=0}$$
$$= <h, A\psi + B\phi + f> \quad \text{as B is symmetric.}$$

Hence $\nabla_\phi L(\phi,\psi) = A\psi + B\phi + f$ (II.7.11)

$$\left[\frac{d}{dt} L(\phi, \psi + tk)\right]_{t=0} = \left[\frac{d}{dt}\left\{<\phi, A(\psi + tk)> + \tfrac{1}{2}<\phi, B\phi>\right.\right.$$
$$\left.\left. - \tfrac{1}{2}<\psi + tk, C(\psi + tk)> + <\phi, f> + <\psi + tk, g>\right\}\right]_{t=0}$$
$$= \left[<\phi, Ak> - \tfrac{1}{2}<\psi, Ck> - \tfrac{1}{2}<k, C\psi> - t<k, k> + <k, g>\right]_{t=0}$$
$$= <k, A^x\phi - C\psi + g> \quad \text{as C is symmetric and A has an adjoint } A^x$$

Hence $\nabla_\psi L(\phi,\psi) = A^x\phi - C\psi + g$ (II.7.12)

This functional will be used extensively in later chapters.

<u>Example 3</u>

Let X be the space $\Phi = \begin{bmatrix}\phi(x)\\ \phi(a)\end{bmatrix}$ with inner product

$$<\Phi_1, \Phi_2>_x = \int_a^b \phi_1(x)\,\phi_2(x)\,dx + \phi_1(a)\,\phi_2(a) \qquad \text{(II.7.13)}$$

Let $L(\phi,\psi) = \int_a^b \phi(x)\,\frac{d\psi(x)}{dx}\,dx + \phi(a)\,\psi(a)$ (II.7.14)

where $\phi(x)$ and $\psi(x)$ are functions which are differentiable and integrable for $x \in (a,b)$

$$\left[\frac{d}{dt} L(\phi + th, \psi)\right]_{t=0} = \left[\frac{d}{dt}\left\{\int_a^b (\phi(x) + th)\,\frac{d\psi(x)}{dx}\,dx\right.\right.$$
$$\left.\left. + (\phi(a) + th)\psi(a)\right\}\right]_{t=0}$$
$$= \left[\int_a^b h\,\frac{d\psi(x)}{dx}\,dx + h\psi(a)\right]_{t=0}$$

$$= \int_a^b h \frac{d\psi(x)}{dx} dx + h\psi(a)$$

Hence $\nabla_\phi L(\phi,\psi) = \begin{pmatrix} \frac{d\psi(x)}{dx} \\ \psi(a) \end{pmatrix}$ (II.7.15)

Using the integration by parts formula,

$$L(\phi,\psi) = \int_a^b -\psi(x) \frac{d\phi(x)}{dx} dx + \phi(b)\psi(b) \qquad \text{(II.7.16)}$$

Let Y be the space $\bar{\Psi} = \begin{bmatrix} \psi(x) \\ \psi(b) \end{bmatrix}$ with inner product

$$\langle \bar{\Psi}_1, \bar{\Psi}_2 \rangle_Y = \int_a^b \psi_1(x)\, \psi_2(x)\, dx + \psi_1(b)\, \psi_2(b) \qquad \text{(II.7.17)}$$

$$\left[\frac{d}{dt} L(\phi, \psi + tk)\right]_{t=0} = \left[\frac{d}{dt}\left\{\int_a^b -(\psi(x) + tk) \frac{d\phi(x)}{dx} dx + \phi(b)(\psi(b) + tk)\right\}\right]_{t=0}$$

$$= \left[-\int_a^b k \frac{d\phi(x)}{dx} dx + k\phi(b)\right]_{t=0}$$

$$= \int_a^b -k \frac{d\phi(x)}{dx} dx + k\phi(b)$$

Hence $\nabla_\psi L(\phi,\psi) = \begin{pmatrix} -\frac{d\phi(x)}{dx} \\ \phi(b) \end{pmatrix}$ (II.7.18)

A similar analysis shows that if $L(\phi,\psi)$ is given by

$$L(\phi,\psi) = \int_a^b \phi(x) \frac{d\psi(x)}{dx} dx - \phi(b)\psi(b) \qquad \text{(II.7.19)}$$

$$= \int_a^b -\psi(x) \frac{d\phi(x)}{dx} dx - \phi(a)\psi(a)$$

then

$$\nabla_\phi L(\phi,\psi) = \begin{pmatrix} \frac{d\psi(x)}{dx} \\ -\psi(b) \end{pmatrix}$$

and $\nabla_\psi L(\phi,\psi) = \begin{pmatrix} -\frac{d\phi(x)}{dx} \\ -\phi(a) \end{pmatrix}$

Note that the notation for the gradients, for example

$$\nabla_{\phi} L(\phi,\psi) = \begin{pmatrix} \dfrac{d\psi(x)}{dx} \\ \psi(a) \end{pmatrix}$$ from equation (II.7.15) means that

$$\nabla_{\phi} L(\phi,\psi) = \frac{d}{dx}\psi \quad \forall\ x \in \ ]a,b[$$ and

$$\nabla_{\phi} L(\phi,\psi) = \psi(a) \quad \text{for } x = a.$$

CHAPTER II

II.8 Convexity and Concavity

Definition II.8.1

A subset of a vector space $C \subseteq Q$ is said to be affine if, for all ϕ_1, $\phi_2 \in Q$,

$$\lambda\phi_1 + (1-\lambda)\phi_2 \in Q \quad \forall \lambda \in \mathbb{R}$$

REFERENCE: (35) p 20

For extremum principles the vector spaces need structure; they must be convex, and functionals defined on them must have convexity or concavity or both.

Definition II.8.2

A subset C of a vector space V is said to be convex if, given $\phi_1, \phi_2 \in C$ then

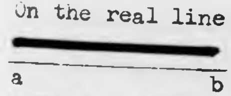
$$\lambda\phi_1 + (1-\lambda)\phi_2 \in C \quad \forall \lambda \in \,]0,1[\qquad \text{(II.8.1)}$$

It can be seen that a convex space is a restricted affine space. In geometrical language, C is convex if whenever it contains two elements it also contains the segment joining these two points.

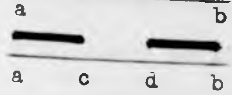
A vector space V is itself both affine and convex.

Examples

On the real line:

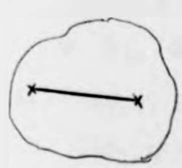


the set $[a, b]$ is convex whereas



the set $[a, c] \cup [d, b]$ is not

In the plane:



This set enclosed by the given boundary is convex,

but this set is not convex.

Definition II.8.3

A functional $L(\phi) : X \to \mathbb{R}$ defined on a convex subset C of X which has a gradient at all points of C, is convex if

$$L(\phi_1) - L(\phi_2) - \langle \phi_1 - \phi_2, \nabla L(\phi_2) \rangle \geq 0 \quad \forall \phi_1, \phi_2 \in C. \qquad (II.8.2)$$

If strict inequality holds in equation (II.8.2) for $\phi_1 \neq \phi_2$ then $L(\phi)$ is strictly convex.

If $L(\phi)$ is (strictly) convex then $-L(\phi)$ is (strictly) concave.

Example

Let $L(\phi) : X \to \mathbb{R}$ be defined by the equation

$$L(\phi) = \langle \phi, B\phi \rangle \qquad (II.8.3)$$

where B is a symmetric, linear operator. Using equation (II.8.2),

$$L(\phi_1) - L(\phi_2) - \langle \phi_1 - \phi_2, \nabla L(\phi_2) \rangle$$
$$= \langle \phi_1, B\phi_1 \rangle - \langle \phi_2, B\phi_2 \rangle - \langle \phi_1 - \phi_2, 2B\phi_2 \rangle$$
$$= \langle \phi_1 - \phi_2, B(\phi_1 - \phi_2) \rangle$$

≥ 0 if B is a positive operator. Hence $L(\phi)$ is a convex functional if B is a positive operator, and is strictly convex if B is a positive-definite operator.

Definition II.8.4

If a functional $L(\phi, \psi) : X \times Y \to \mathbb{R}$, which is defined on a convex subset C of X, has a gradient with respect to ϕ at all points of C, then

(i) $L(\phi,\psi)$ is said to be convex with respect to ϕ for all $\psi \in Y$ if

$$L(\phi_1, \psi) - L(\phi_2, \psi) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2, \psi) \rangle \geq 0 \quad \forall \phi_1, \phi_2 \in C \text{ and } \forall \psi \in Y \qquad (II.8.4)$$

(ii) $L(\phi,\psi)$ is said to be concave with respect to ϕ for all $\psi \in Y$ if

$$L(\phi_1, \psi) - L(\phi_2, \psi) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_1, \psi) \rangle \geq 0 \quad \forall \phi_1, \phi_2 \in C \text{ and } \forall \psi \in Y \qquad (II.8.5)$$

$L(\phi,\psi)$ is <u>strictly convex</u> (<u>strictly concave</u>) if strict inequality holds in equation (II.8.4) ((II.8.5)) for all $\phi_1 \neq \phi_2$.

<u>Definition II.8.5</u>

If a functional $L(\phi,\psi) : X \times Y \to \mathbb{R}$, which is defined on a convex subset D of Y, has a gradient with respect to ψ at all points of D, then

(i) $L(\phi,\psi)$ is said to be <u>convex</u> with respect to ψ for all $\phi \in X$ if

$$L(\phi, \psi_1) - L(\phi, \psi_2) - \langle \psi_1 - \psi_2, \nabla_\psi L(\phi, \psi_2)\rangle \geq 0 \qquad \text{(II.8.6)}$$

$\forall\ \phi \in X$ and $\forall\ \psi_1, \psi_2 \in D$

(ii) $L(\phi,\psi)$ is said to be <u>concave</u> with respect to ψ for all $\phi \in X$ if

$$L(\phi, \psi_1) - L(\phi, \psi_2) - \langle \psi_1 - \psi_2, \nabla_\psi L(\phi, \psi_1)\rangle \geq 0 \qquad \text{(II.8.7)}$$

$\forall\ \phi \in X$ and $\forall\ \psi_1, \psi_2 \in D$.

As before, $L(\phi,\psi)$ is strictly convex (strictly concave) if strict inequality holds in equation (II.8.6) ((II.8.7)) for all $\psi_1 \neq \psi_2$.

REFERENCES: (36) pp 3-4; (9) pp 29-32

<u>Example</u>

Let
$$L(\phi,\psi) = \langle \phi, A\psi\rangle + \tfrac{1}{2}\langle \phi, B\phi\rangle - \tfrac{1}{2}\langle \psi, C\psi\rangle + \langle f,\phi\rangle + \langle g,\psi\rangle \qquad \text{(II.8.8)}$$

where the linear operator A has an adjoint A^x such that $\langle \phi, A\psi\rangle = \langle \psi, A^x\phi\rangle$, and B and C are symmetric linear operators.

Applying equations (II.8.4) to (II.8.7) in turn; and noting that

$\nabla_\phi L(\phi,\psi) = A\psi + B\phi + f$, $\nabla_\psi L(\phi,\psi) = A^x\phi - C\psi + g$,

$$L(\phi_1, \psi) - L(\phi_2, \psi) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2, \psi)\rangle$$
$$= \tfrac{1}{2}\langle \phi_1 - \phi_2, B(\phi_1 - \phi_2)\rangle \qquad \text{(II.8.9)}$$

$$L(\phi_1, \psi) - L(\phi_2, \psi) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_1, \psi)\rangle$$
$$= -\tfrac{1}{2}\langle \phi_1 - \phi_2, B(\phi_1 - \phi_2)\rangle \qquad \text{(II.8.10)}$$

$$L(\phi, \psi_1) - L(\phi, \psi_2) - \langle \psi_1 - \psi_2, \nabla_\psi L(\phi, \psi_2)\rangle$$
$$= -\tfrac{1}{2}\langle \psi_1 - \psi_2, C(\psi_1 - \psi_2)\rangle \qquad \text{(II.8.11)}$$

$$L(\phi,\psi_1) - L(\phi,\psi_2) - \langle \psi_1 - \psi_2, \nabla_\psi L(\phi,\psi_1)\rangle$$
$$= \tfrac{1}{2}\langle \psi_1 - \psi_2, C(\psi_1 - \psi_2)\rangle \qquad \text{(II.8.12)}$$

Hence $L(\phi,\psi)$ is convex in ϕ if B is a positive operator, concave in ϕ if B is a negative operator, convex in ψ if C is a negative operator and concave in ψ if C is a positive operator.

<u>Definition II.8.6</u>

If $L : X \times Y \to \mathbb{R}$ has a gradient with respect to ϕ at all points of $C \subseteq X$ and a gradient with respect to ψ at all points of $D \subseteq Y$ then $L(\phi,\psi)$ is <u>Jointly Convex</u> in ϕ and ψ if, for all pairs (ϕ_1,ψ_1) and $(\phi_2,\psi_2) \in C \times D$,

$$-L(\phi_1,\psi_1) + L(\phi_2,\psi_2) + \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_1,\psi_1)\rangle$$
$$+ \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_1,\psi_1)\rangle \geq 0 \qquad \text{(II.8.13)}$$

Similarly, $L(\phi,\psi)$ is said to be <u>Jointly Concave</u> in ϕ and ψ if, for all pairs (ϕ_1,ψ_1) and $(\phi_2,\psi_2) \in C \times D$,

$$-L(\phi_1,\psi_1) + L(\phi_2,\psi_2) + \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2,\psi_2)\rangle$$
$$+ \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_2,\psi_2)\rangle \geq 0 \qquad \text{(II.8.14)}$$

The joint convexity (joint concavity) is strict if the inequality in equation (II.8.13) ((II.8.14)) is strict.

It should also be noted that joint convexity (joint concavity) in ϕ and ψ implies separate convexity (concavity) in ϕ and ψ, but not conversely.

REFERENCES: (68) p 187; (60) p 580

<u>Example 1</u>

Let $L(\phi,\psi) = \langle \phi, A\psi\rangle \equiv \langle \psi, A^x \phi\rangle \qquad \text{(II.8.15)}$

$$-L(\phi_1,\psi_1) + L(\phi_2,\psi_2) + \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_1,\psi_1)\rangle$$
$$+ \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_1,\psi_1)\rangle$$
$$= \langle \phi_1 - \phi_2, A(\psi_1 - \psi_2)\rangle \qquad \text{(II.8.16)}$$

and

$$-L(\phi_1, \psi_1) + L(\phi_2, \psi_2) + \langle \phi_1 - \phi_2, \nabla_\phi \; L(\phi_2, \psi_2) \rangle$$
$$+ \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_2, \psi_2) \rangle$$
$$= -\langle \phi_1 - \phi_2, A\,(\psi_1 - \psi_2) \rangle \qquad \text{(II.8.17)}$$

For this functional, the application of equation (II.8.13) and (II.8.14) result in an expression containing both ϕ and ψ terms, for which we cannot say anything about the signs, and thus we cannot show that the functional is jointly convex or jointly concave; hence a mixed term such as $\langle \phi, A\psi \rangle$ cannot be included in any functional that is required to be jointly convex or jointly concave.

<u>Example 2</u>

Let $L(\phi, \psi) = \frac{1}{2} \langle \phi, B\phi \rangle + \frac{1}{2} \langle \psi, C\psi \rangle + \langle \phi, f \rangle \; + \langle \psi', g \rangle$ (II.8.18)

where B and C are linear symmetric operators.

$$-L(\phi_1, \psi_1) + L(\phi_2, \psi_2) + \langle \phi_1 - \phi_2, \nabla_\phi \; L(\phi_1, \psi_1) \rangle$$
$$+ \langle \phi_1 - \phi_2, \nabla_\phi \; L(\phi_1, \psi_1) \rangle$$
$$= \tfrac{1}{2} \langle \phi_1 - \phi_2, B(\phi_1 - \phi_2) \rangle \; + \tfrac{1}{2} \langle \psi_1 - \psi_2, C\,(\psi_1 - \psi_2) \rangle \qquad \text{(II.8.19)}$$

$$-L(\phi_1, \psi_1) + L(\phi_2, \psi_2) + \langle \phi_1 - \phi_2, \nabla_\phi \; L(\phi_2, \psi_2) \rangle$$
$$+ \langle \psi_1 - \psi_2, \nabla_\psi \; L(\phi_2, \psi_2) \rangle$$
$$= -\tfrac{1}{2} \langle \phi_1 - \phi_2, B(\phi_1 - \phi_2) \rangle \; - \tfrac{1}{2} \langle \psi_1 - \psi_2, C(\psi_1 - \psi_2) \rangle \qquad \text{(II.8.20)}$$

Hence the functional given by equation (II.8.18) is jointly convex if B and C are positive operators and jointly concave if B and C are negative operators.

II.9 Saddle Functionals

Definition II.9.1

A functional $L(\phi,\psi)$ $X \times Y \rightarrow \mathbb{R}$ defined on a convex subset C of X and a convex subset D of Y is called a convex/concave saddle functional if $L(\phi,\psi)$ is convex in ϕ for all ϕ in C and each fixed $\psi \in Y$ and concave in ψ for all ψ in D and each fixed $\phi \in X$.

If $L(\phi,\psi)$ is a convex/concave saddle functional then $- L(\phi,\psi)$ is a concave/convex saddle functional, Figure (II.9.1) shows a schematic view of a convex-concave saddle functional $L(\phi,\psi)$.

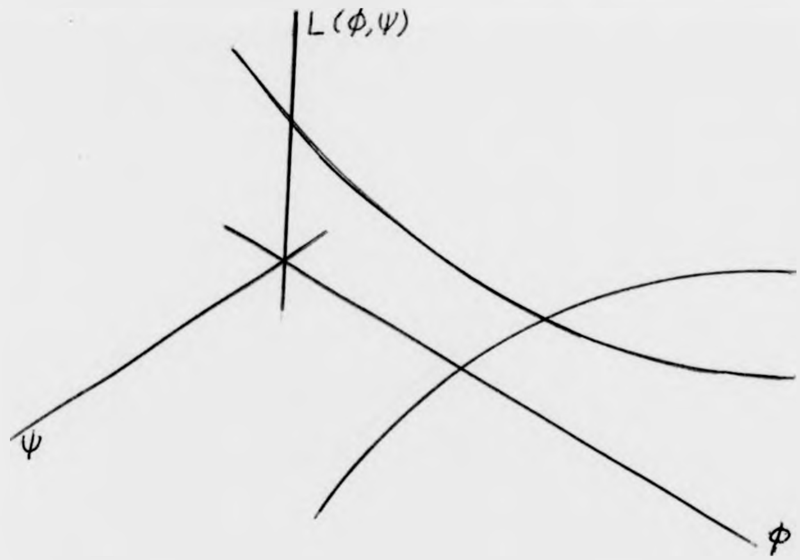


Figure (II.9.1)

Theorem II.9.1

Let $L(\phi,\psi)$: $X \times Y \rightarrow \mathbb{R}$ have a gradient with respect to ϕ at all points of the convex subset C of X and a gradient with respect to ψ at all points of the convex subset D of Y. Then the statement '$L(\phi,\psi)$ is a convex/concave saddle functional' is equivalent to

$$L(\phi_1, \psi_1) - L(\phi_2, \psi_2) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2, \psi_2) \rangle$$
$$- \langle \psi_1 - \psi_2, \nabla_\phi L(\phi_1, \psi_1) \rangle \geq 0$$
$$\forall \phi_1, \phi_2 \in C \text{ and } \forall \psi_1, \psi_2 \in D \qquad \text{(II.9.1)}$$

Proof

$L(\phi,\psi)$ is a convex/concave saddle functional if it is convex with respect to ϕ for all $\psi \in Y$ and concave with respect to ψ for all $\phi \in X$. By equations (II.8.4) and (II.8.7), $L(\phi,\psi)$ is a convex/concave saddle functional if

$$L(\phi_1,\psi) - L(\phi_2,\psi) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2,\psi)\rangle \geq 0 \qquad \text{(II.9.2)}$$
$$\forall \phi_1, \phi_2 \in C \text{ and } \forall \psi \in Y$$

and

$$L(\phi,\psi_1) - L(\phi,\psi_2) - \langle \psi_1 - \psi_2, \nabla_\psi L(\phi,\psi_1)\rangle \geq 0 \qquad \text{(II.9.3)}$$
$$\forall \phi \in X \text{ and } \forall \psi_1, \psi_2 \in D.$$

Let $\psi = \psi_2$ in equation (II.9.2) and $\phi = \phi_1$ in equation (II.9.3), and add:

$$L(\phi_1,\psi_2) - L(\phi_2,\psi_2) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2,\psi_2)\rangle$$
$$+ L(\phi_1,\psi_1) - L(\phi_1,\psi_2) - \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_1,\psi_1)\rangle \geq 0$$

or $$L(\phi_1,\psi_1) - L(\phi_2,\psi_2) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2,\psi_2)\rangle$$
$$- \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_1,\psi_1)\rangle \geq 0$$

as required.

If strict inequality holds in equation (II.9.1) for all $\phi_1 \neq \phi_2$ and for all $\psi_1 \neq \psi_2$ then $L(\phi,\psi)$ is a strict convex-concave saddle functional.

Similarly, let $L : X \times Y \to \mathbb{R}$ have a gradient with respect to ϕ at all points of the convex subset C of X and a gradient with respect to ψ at all points of the convex subset D of Y. Then the statement '$L(\phi,\psi)$ is a concave-convex saddle functional' is equivalent to

$$L(\phi_1,\psi_1) - L(\phi_2,\psi_2) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_1,\psi_1)\rangle$$
$$- \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_2,\phi_2)\rangle \geq 0$$
$$\forall \phi_1, \phi_2 \in C \text{ and } \forall \psi_1, \psi_2 \in D \qquad \text{(II.9.4)}$$

As before, if strict inequality holds in equation (II.9.4) for all $\phi_1 \neq \phi_2$

Hence $L(\phi,\psi)$ is a concave-convex saddle functional if B and C are both negative operators, and a strict concave-convex saddle functional if one of B or C is negative-definite.

CHAPTER II

II.10 Variational Principles

Problems which are concerned with finding the stationary values of a functional are called variational problems. Extrema, where they exist, occur at the stationary points of the functional. Non-stationary extrema are often described as optimization problems but the distinction is not rigid.

REFERENCE: (37) p 10

Definition II.10.1

The functional $L(\phi) : X \to \mathbb{R}$ is said to be stationary at $\phi = \phi_e$ if

$$\nabla L(\phi_e) = 0 \qquad \text{(II.10.1)}$$

(assuming, of course, that the gradient exists.)

REFERENCE: (9) p 25

A variational principle does not lead automatically to an extremum principle without further conditions. These can be provided by convexity and concavity, although it is possible to have, for example, a minimum of a non-convex functional. An example is provided by the functional $L(\phi) = ||\phi||^{\frac{1}{2}}$ where $\phi \in X_1$, a normed vector space. This has a minimum at $\phi = 0$ but is not stationary nor convex in ϕ.

Lemma II.10.1

Let $L(\phi)$ be a differentiable convex functional on C, a convex subset of X. If $L(\phi)$ has a stationary point at ϕ_e, that is $\nabla L(\phi_e) = 0$, then

$$L(\phi) \geq L(\phi_e) \quad \forall \phi \in C \qquad \text{(II.10.2)}$$

This is a minimum principle.

Proof

By equation (II.8.2), as $L(\phi)$ is a differentiable convex functional,

$$L(\phi_1) - L(\phi_2) - \langle \phi_1 - \phi_2, \nabla L(\phi_2) \rangle \geq 0$$
$$\forall \phi_1, \phi_2 \in C$$

Take $\phi_2 = \phi_e$ and $\phi_1 = \phi$

Then $L(\phi) - L(\phi_e) - \langle \phi - \phi_e, \nabla L(\phi_e) \rangle \geq 0$

But $\nabla L(\phi_e) = 0$

Hence $L(\phi) - L(\phi_e) \geq 0$

and so $L(\phi) \geq L(\phi_e)$

<u>Lemma (II.10.2)</u>

Let $L(\psi)$ be a differentiable concave functional on D, a concave subset of Y. If $L(\psi)$ has a stationary point at ψ_e, that is $\nabla L(\psi_e) = 0$, then

$$L(\psi) \leq L(\psi_e) \quad \forall \psi \in D \qquad \text{(II.10.3)}$$

This is a <u>maximum</u> principle.

REFERENCE: (9) p 31

<u>Example</u>

Let $L(\phi) = \frac{1}{2}\langle \phi, A\phi \rangle - \langle \phi, f \rangle \quad \phi \in C$

where A is a symmetric, positive operator.

$\nabla L(\phi) = A\phi - f$

Then $L(\phi)$ has a stationary point at ϕ_e where $A\phi_e = f$.

Using equation (II.8.2),

$L(\phi_1) - L(\phi_2) - \langle \phi_1 - \phi_2, \nabla L(\phi_2) \rangle$

$= \frac{1}{2}\langle \phi_1 - \phi_2, A(\phi_1 - \phi_2) \rangle$

≥ 0 as A is a positive operator; thus $L(\phi)$ is a convex functional.

$$\begin{aligned} L(\phi_e) &= \langle \tfrac{1}{2}\phi_e, A\phi_e \rangle - \langle \phi_e, f \rangle \\ &= -\tfrac{1}{2}\langle \phi_e, f \rangle \end{aligned}$$

Hence $\langle \phi, A\phi \rangle - \langle \phi, f \rangle \geq -\frac{1}{2}\langle \phi_e, f \rangle \ \forall \phi \in C$

CHAPTER II

II.11 Dual Variational Principles

Techniques concerned with finding the stationary values of a functional of two variables defined on the product space of two vector spaces are called dual variational principles.

Definition II.11.1

Let $L(\phi,\psi) : X \times Y \to \mathbb{R}$ be a functional for which the functional derivatives $\nabla_\phi L(\phi,\psi)$ and $\nabla_\psi L(\phi,\psi)$ exist. Then $L(\phi,\psi)$ is stationary at $\phi = \phi_e$ and $\psi = \psi_e$ if (ϕ_e, ψ_e) are solutions of the pair of equations

$\nabla_\phi L(\phi_e, \psi_e) = 0$ and $\nabla_\psi L(\phi_e, \psi_e) = 0$

REFERENCE: (9) p 25

Example

Let
$$L(\phi,\psi) = \langle \phi, A\psi \rangle + \tfrac{1}{2}\langle \phi, B\phi \rangle - \tfrac{1}{2}\langle \psi, C\psi \rangle + \langle f, \phi \rangle + \langle g, \psi \rangle \qquad \text{(II.11.1)}$$

where A is a linear operator with an adjoint A^x such that $\langle \phi, A\psi \rangle = \langle \psi, A^x\phi \rangle$ and B and C are symmetric linear operators. From example 2 of section II.6, the gradients are given by

$$\nabla_\phi L(\phi,\psi) = A\psi + B\phi + f \qquad \text{(II.11.2)}$$

$$\nabla_\psi L(\phi,\psi) = A^x\phi - C\psi + g \qquad \text{(II.11.3)}$$

Hence the stationary value $L(\phi_e, \psi_e)$ is given by equation (II.11.1) with (ϕ_e, ψ_e) satisfying

$$A\psi_e + B\phi_e + f = 0 \qquad \text{(II.11.4)}$$

and
$$A^x\phi_e - C\psi_e + g = 0 \qquad \text{(II.11.5)}$$

(assuming such a solution exists). Substituting these two equations into

$$L(\phi_e, \psi_e) = \langle \phi_e, A\psi_e \rangle + \tfrac{1}{2}\langle \phi_e, B\phi_e \rangle - \tfrac{1}{2}\langle \psi_e, C\psi_e \rangle + \langle f, \phi_e \rangle + \langle g, \psi_e \rangle$$

gives

$$L(\phi_e, \psi_e) = \tfrac{1}{2}\langle \phi_e, f \rangle + \tfrac{1}{2}\langle \psi_e, g \rangle \qquad \text{(II.11.6)}$$

CHAPTER II

II.12 Dual Extremum Principles

As in the one-dimensional case (see section II.10), dual variational principles do not lead automatically to dual extremum principles unless further conditions are placed on the functional $L(\phi,\psi)$.

As we shall see, sufficient conditions are that L is convex with respect to one variable and concave with respect to the other variable - that is $L(\phi,\psi)$ is a saddle functional.

The following theorem is the central result in dual extremum principles.

Theorem II.12.1

Let $L(\phi,\psi) : X \times Y \to \mathbb{R}$ be a convex-concave saddle functional with a stationary value at (ϕ_e, ψ_e). Then

$$L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha) \qquad \text{(II.12.1)}$$

where

(ϕ_β, ψ_β) is any pair which satisfies $\nabla_\phi L(\phi, \psi) = 0$

$(\phi_\alpha, \psi_\alpha)$ is any pair which satisfies $\nabla_\psi L(\phi, \psi) = 0$

and (ϕ_e, ψ_e) satisfies both $\nabla_\phi L(\phi,\psi) = 0$ and $\nabla_\psi L(\phi,\psi) = 0$.

REFERENCE: (9) p 38

Proof

By equation (II.9.1) as $L(\phi,\psi)$ is a convex-concave saddle functional,

$$L(\phi_1,\psi_1) - L(\phi_2, \psi_2) - \langle \phi_1 - \phi_2 , \nabla_\phi L(\phi_2, \psi_2)\rangle$$
$$- \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_1, \psi_1)\rangle \geq 0$$

$$\forall\ \phi_1, \phi_2 \in C \subseteq X \text{ and } \forall\ \psi_1, \psi_2 \in D \subseteq Y \qquad \text{(II.12.2)}$$

Lower Bound

Take $(\phi_1, \psi_1) = (\phi_e, \psi_e)$ and $(\phi_2, \psi_2) = (\phi_\beta, \psi_\beta)$ Then equation (II.12.2) becomes

$$L(\phi_e, \psi_e) - L(\phi_\beta, \psi_\beta) - \langle \phi_e - \phi_\beta, \nabla_\phi L(\phi_\beta, \psi_\beta)\rangle$$
$$- \langle \psi_e - \psi_\beta, \nabla_\psi L(\phi_e, \psi_e)\rangle \geq 0 \qquad \text{(II.12.3)}$$

By the conditions of the theorem, $\nabla_\phi L(\phi_\beta, \psi_\beta) = 0$ and $\nabla_\psi L(\phi_e, \phi_e) = 0$. Hence equation (II.12.3) reduces to

$$L(\phi_e, \psi_e) - L(\phi_\beta, \psi_\beta) \geqslant 0 \text{ or}$$

$$L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \qquad \text{(II.12.4)}$$

<u>Upper Bound</u>

Take $(\phi_1, \psi_1) = (\phi_\alpha, \psi_\alpha)$ and $(\phi_2, \psi_2) = (\phi_e, \psi_e)$ then equation (II.12.2) becomes

$$L(\phi_\alpha, \psi_\alpha) - L(\phi_e, \psi_e) - \langle \phi_\alpha - \phi_e, \nabla_\phi L(\phi_e, \psi_e) \rangle$$
$$- \langle \psi_\alpha - \psi_e, \nabla_\psi L(\phi_\alpha, \psi_\alpha) \rangle \geqslant 0 \qquad \text{(II.12.5)}$$

As $\nabla_\phi L(\phi_e, \psi_e) = 0$ and $\nabla_\psi L(\phi_\alpha, \psi_\alpha) = 0$, equation (II.2.5) reduces to

$$L(\phi_\alpha, \psi_\alpha) - L(\phi_e, \psi_e) \geqslant 0 \text{ or}$$

$$L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha) \qquad \text{(II.12.6)}$$

Putting equations (II.12.4) and (II.12.6) together gives

$L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\beta, \psi_\beta)$ as required.

Figure (II.12.1) shows a geometrical interpretation of the dual extremum principles for a convex-concave saddle functional $L(\phi, \psi)$.

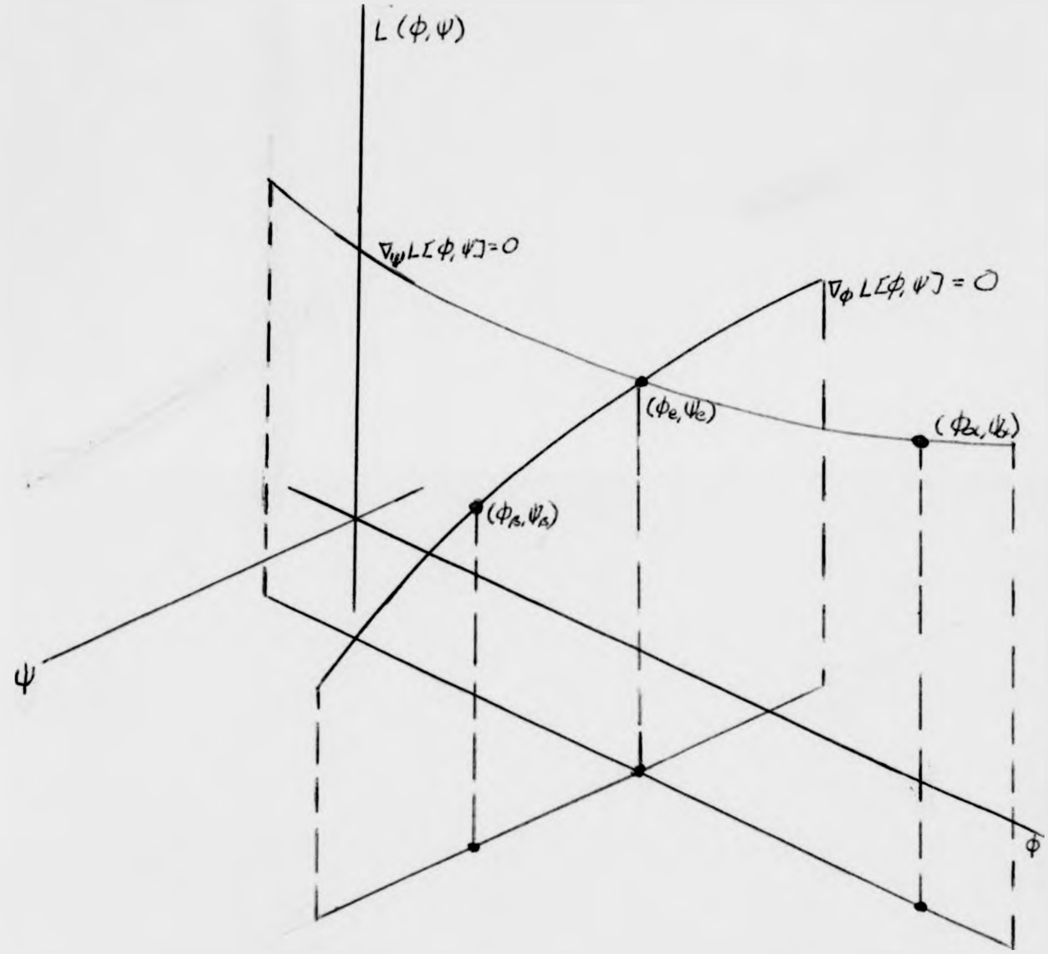


Figure II.12.1

Example

Let $L(\phi,\psi) : X \times Y \to \mathbb{R}$ be defined by

$$L(\phi,\psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B\phi\rangle - \tfrac{1}{2}\langle\psi, C\psi\rangle$$
$$+ \langle f,\phi\rangle + \langle g,\psi\rangle \qquad \text{(II.12.7)}$$

where A is a linear operator with an adjoint A^x such that $\langle\phi, A\psi\rangle = \langle\psi, A^x\phi\rangle$, and B and C are linear, symmetric positive-definite operators; then from the example in section II.9, $L(\phi,\psi)$ is a convex-concave saddle functional. It's gradients are given by

$$\nabla_{\varphi} L(\phi,\psi) = A\psi + B\phi + f \qquad \text{(II.12.8)}$$

$$\nabla_{\psi} L(\phi,\psi) = A^{x}\phi - C\psi + g \qquad \text{(II.12.9)}$$

By theorem (II.12.1), the dual extremum principles are:

$$L(\phi_{\beta},\psi_{\beta}) \leq L(\phi_{e},\phi_{e}) \leq L(\phi_{\alpha},\psi_{\alpha})$$

where $A\psi + B\phi + f = 0,$ (II.12.10)

$$A^{x}\phi - C\psi + g = 0 \qquad \text{(II.12.11)}$$

and $A\psi_{e} + B\phi_{e} + f = 0$ and $A^{x}\phi_{e} - C\psi_{e} + g = 0$ (II.12.12)

Substituting equations (II.12.10) - (II.12.12) into equation (II.12.7) gives

$$L(\phi_{\beta},\psi_{\beta}) = -\tfrac{1}{2}\langle\phi_{\beta}, B\phi_{\beta}\rangle - \tfrac{1}{2}\langle\psi_{\beta}, C\psi_{\beta}\rangle + \langle\psi_{\beta}, g\rangle \qquad \text{(II.12.13)}$$

$$L(\phi_{\alpha},\psi_{\alpha}) = \tfrac{1}{2}\langle\phi_{\alpha}, B\phi_{\alpha}\rangle + \tfrac{1}{2}\langle\psi_{\alpha}, C\psi_{\alpha}\rangle + \langle\phi_{\alpha}, f\rangle \qquad \text{(II.12.14)}$$

$$L(\phi_{e},\psi_{e}) = \tfrac{1}{2}\langle\phi_{e}, f\rangle + \tfrac{1}{2}\langle\psi_{e}, g\rangle \qquad \text{(II.12.15)}$$

These dual extremum principles will be used extensively in the following chapters.

CHAPTER II

II.13 Uniqueness

Theorem II.13.1

If $L(\phi,\psi) : X \times Y \to \mathbb{R}$ is a strict saddle functional, and the stationary point (ϕ_e,ψ_e) exists, then it is unique.

Proof

By theorem (II.9.1), $L(\phi,\psi) : X \times Y \to \mathbb{R}$ is a strict saddle functional if

$$L(\phi_1,\psi_1) - L(\phi_2,\psi_2) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2,\psi_2)\rangle$$
$$- \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_1,\psi_1)\rangle > 0$$
$$\forall \phi_1, \phi_2 \in C \subseteq X \text{ and } \forall \psi_1, \psi_2 \in D \subseteq Y \qquad \text{(II.13.1)}$$

As (ϕ_1, ϕ_2) and (ψ_1, ψ_2) are arbitrary, we can reverse the labelling in equation (II.13.1) to give

$$L(\phi_2,\psi_2) - L(\phi_1,\psi_1) + \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_1,\psi_1)\rangle$$
$$+ \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_2,\psi_2)\rangle > 0 \qquad \text{(II.13.2)}$$

Adding equations (II.13.1) and (II.13.2) results in

$$-\langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2,\psi_2) - \nabla_\phi L(\phi_1,\psi_1)\rangle$$
$$+\langle \psi_1 - \psi_2, \nabla_\psi L(\phi_2,\psi_2) - \nabla_\psi L(\phi_1,\psi_1)\rangle > 0 \qquad \text{(II.13.3)}$$

From theorem (II.12.1), a stationary value of $L(\phi,\psi)$ satisfies $\nabla_\phi L(\phi,\psi) = 0$ and $\nabla_\psi L(\phi,\psi) = 0$. Assume that there are two distinct stationary values (ϕ_1, ψ_1) and (ϕ_2, ψ_2); then

$$\nabla_\phi L(\phi_1,\psi_1) = \nabla_\phi L(\phi_2,\psi_2) = \nabla_\psi L(\phi_1,\psi_1) = \nabla_\psi L(\phi_2,\psi_2) = 0 \qquad \text{(II.13.4)}$$

Inserting the last equations into equation (II.13.3) gives

$$0 > 0$$

which is a contradiction. Hence $(\phi_1,\psi_1) = (\phi_2,\psi_2)$ and the stationary point is unique.

REFERENCE: (59) pp 314-317

If the saddle property is only partly strict, say strictly convex in ϕ but only weakly concave in ψ, then ϕ_e, where it exists, is unique but nothing can be concluded about the uniqueness of ψ_e.

REFERENCE: (50) p 146

Example

In section II.9, it was shown that the saddle functional given by equation (II.12.8) in the previous section is a strict convex/concave saddle functional; hence the stationary point (ϕ_e, ψ_e), specified by equations (II.12.13), is unique (assuming that it exists).

The remaining sections of this chapter give various results which will be used in the following chapters. Section II.14 shows how the boundary conditions associated with a second order differential equation can be incorporated into a quadratic functional whose gradients represent the equation; section II.15 considers integral operators, and sections II.16 to II.18 give some convergence results.

CHAPTER II

II.14 Boundary Conditions for Second Order Ordinary Differential Equations

The purpose of this section is to specify a quadratic functional $L(\phi, \psi)$ whose gradients, when set to zero, give the problem

$$\frac{d}{dt}\left[r(t) \frac{d\phi(t)}{dt} \right] + p(t)\ \phi(t) + q\ (t) = 0,\ t \in \left[a,b\right],$$

$$A_1\ \phi(a) + A_2\ \phi'(a) = A_3 \qquad \text{(II.14.1)}$$

$$B_1\ \phi(b) + B_2\ \phi'(b) = B_3$$

where $p(t)$, $q(t)$ and $r(t)$ are real functions of t, with $r(t) \neq 0$ on $[a,b]$; and A_1, A_2, A_3, B_1, B_2 and B_3 are real numbers such that A_1 and A_2 are not both zero and B_1 and B_2 are not both zero.

This section is intended as a summary of the part of section 9 of (50) which is concerned with boundary conditions for second order differential equations, and will be used later in this thesis.

We will denote the values Φ and Ψ by

$$\Phi = \begin{bmatrix} \phi(t) \\ \phi(b) \\ \phi(a) \end{bmatrix} \quad \text{and} \quad \Psi = \begin{bmatrix} \psi(t) \\ \psi(b) \\ \psi(a) \end{bmatrix} \qquad \text{(II.14.2)}$$

where $\phi(t)$ and $\psi(t)$ are functions which are differentiable and integrable for $x \in [a,b]$.

We then have the inner product

$$\langle r,s \rangle = \int_a^b r(t)\ s(t)\ dt + r(b)\ s(b) + r(a)\ s(a) \qquad \text{(II.14.3)}$$

(a) The general boundary conditions in equation (II.14.1) will result from setting $\nabla_{\Phi} L(\Phi,\Psi) = 0$ and $\nabla_{\Psi} L(\Phi,\Psi) = 0$ if we define $L(\Phi,\Psi)$ as:

$$L(\Phi,\Psi) = \int_a^b \left\{ \phi(t) \frac{d\psi(t)}{dt} + \tfrac{1}{2}\, p(t)\ (\phi(t))^2 + \frac{1}{2r(t)}\ (\psi(t))^2 + \phi(t)\ q(t) \right\} dt$$

$$- \phi(b)\ \psi(b) - \tfrac{1}{2} \frac{B_1}{B_2}\, r(b)\ (\phi(b))^2 + \frac{B_3}{B_2}\ \phi(b)\ r(b)$$

$$- \frac{A_2\ (\psi(a))^2}{2A_1\ r(a)} + \frac{A_3}{A_1}\psi(a) \qquad \text{(II.14.4)}$$

Using the derivatives given in example 3 after definition (II.7.2), we can find $\nabla_{\Phi} L(\Phi,\Psi)$ and $\nabla_{\Psi} L(\Phi,\Psi)$: these are

$$\nabla_{\Phi} L(\Phi,\Psi) = \begin{pmatrix} \frac{d\psi(t)}{dt} + p(t)\ \phi(t) + q(t) \\ -\ \psi(b) - \frac{B_1}{B_2}\ r(b)\,\phi(b) + \frac{B_3}{B_2}\ r(b) \\ 0 \end{pmatrix} \tag{II.14.5}$$

$$\nabla_{\Psi} L(\Phi,\Psi) = \begin{pmatrix} -\frac{d\phi(t)}{dt} + \frac{\psi(t)}{r(t)} \\ 0 \\ -\phi(a) - \frac{A_2}{A_1}\ \frac{\psi(a)}{r(a)} + \frac{A_3}{A_1} \end{pmatrix} \tag{II.14.6}$$

Setting these two derivatives equal to zero gives the problem given in equation (II.14.1), as required. Obviously we cannot have A_1 or B_2 equal to zero.

It can easily be shown, using theorem (II.9.1), that $L(\Phi,\Psi)$ is a convex/concave saddle functional if:

$$\left.\begin{array}{l} p(t) \geq 0\ \forall\ t \in [a,b],\ r(t) < 0\ \forall\ t \in [a,b] \\ \frac{B_1}{B_2} \geq 0 \text{ and } \frac{A_2}{A_1} \leq 0 \end{array}\right\} \tag{II.14.7}$$

and $L(\Phi,\Psi)$ is a concave/convex saddle functional if

$$\left.\begin{array}{l} p(t) \leq 0\ \forall\ t \in [a,b],\ r(t) > 0\ \forall\ t \in [a,b] \\ \frac{B_1}{B_2} \leq 0 \text{ and } \frac{A_2}{A_1} \geq 0 \end{array}\right\} \tag{II.14.8}$$

(b) Less general boundary conditions can be obtained if we specify $L(\Phi,\Psi)$ as:

$$\begin{aligned} L(\Phi,\Psi) = \int_a^b \Big\{ & \phi(t)\ \frac{d\psi(t)}{dt} + \tfrac{1}{2}\, p(t)\ (\phi(t))^2 + \frac{1}{2\, r(t)}\ (\psi(t))^2 \\ & + \phi(t)\ q(t) \Big\}\ dt \\ & + c\,\phi(b)\ \psi(b) + e r(b)\phi(b) + h\,\psi(b) \\ & + d\,\phi(a)\ \psi(a) + j r(a)\phi(a) + k\,\psi(a) \end{aligned}$$

where $c, d, e, h, j, k \in \mathbb{R}$ (II.14.9)

For this functional, the derivatives are

$$\nabla_{\bar{\Phi}} \; L(\bar{\Phi},\bar{\Psi}) = \begin{pmatrix} \dfrac{d\Psi(t)}{dt} + p(t)\,\phi(t) + q(t) \\ c\,\Psi(b) + er(b) \\ d\,\Psi(a) + jr(a) \end{pmatrix} \qquad \text{(II.14.10)}$$

$$\nabla_{\bar{\Psi}} \; L(\bar{\Phi},\bar{\Psi}) = \begin{pmatrix} -\dfrac{d\Phi(t)}{dt} + \dfrac{\Psi(t)}{r(t)} \\ (c+1)\,\phi(b) + h \\ (d-1)\,\phi(a) + k \end{pmatrix} \qquad \text{(II.14.11)}$$

The following table details the values the numbers c, d, e, h, j and k must take for the various boundary conditions:

c	d	e	h	j	k	Boundary Conditions	
0	0	0	-n	0	m	$\phi(a) = m$	$\phi(b) = n$
-1	1	n	0	-m	0	$\phi'(a) = m$	$\phi'(b) = n$
-1	0	n	0	0	m	$\phi(a) = m$	$\phi'(b) = n$
0	1	0	-n	-m	0	$\phi'(a) = m$	$\phi(b) = n$

Table II.14.1

Using theorem (II.9.1) again, we can show that $L(\bar{\Phi},\bar{\Psi})$ is a convex/concave saddle functional if $p(t) \geq 0$ and $r(t) < 0$ and a concave/convex saddle functional if $p(t) \leq 0$ and $r(t) > 0$.

CHAPTER II

II.15 Integral Operators

In several of the later sections we will need a linear, self-adjoint operator to illustrate the methods developed in this thesis. As linear, symmetric integral operators are bounded (51), p 225, they will be discussed in some detail here.

The Fredholm integral equation of the second kind takes the form

$$\phi(x) - \lambda\int_a^b k(x,y)\,\phi(y)\,dy + f(x) = 0 \qquad \text{(II.15.1)}$$

where $f(x)$ is a given function defined at every point x of the real set $[a,b]$, $\phi(x)$ is an unknown function in $[a,b]$ and the kernel $k(x,y)$ is a given function defined for every pair of points x and y of $[a,b]$, λ is a real number. Only real functions $\phi(x)$ and $k(x,y)$ will be considered.

Equation (II.15.1) can be written as

$$(I - \lambda K)\,\phi(x) = -f(x) \qquad \text{(II.15.2)}$$

where $$K\phi(x) = \int_R k(x,y)\,\phi(y)\,dy \qquad \text{(II.15.3)}$$

In order to find dual extremum principles for the problem specified by the above two equations, a saddle functional $L(\phi,\psi)$ is required whose gradients, when set to zero, give equation (II.15.2). Starting from the general quadratic functional considered earlier,

$$L(\phi,\psi) = \langle\phi,\, A\psi\rangle + \tfrac{1}{2}\langle\phi,\, B\phi\rangle - \tfrac{1}{2}\langle\psi,\, C\psi\rangle + \langle\phi,\, f\rangle + \langle\psi,\, g\rangle \qquad \text{(II.15.4)}$$

the operators A, B, C and functions f and g need to be specified to give equation (II.15.2), with the proviso that the linear operator A has an adjoint A^x such that $\langle\phi,\, A\psi\rangle = \langle\psi,\, A^x\phi\rangle$, and the linear, symmetric operators B and C are both positive-definite or both negative-definite, to ensure that $L(\phi,\psi)$ is a strict saddle functional. This specification is not unique; there are various ways in which A, B, C, f and g can be chosen. The following equations provide the simplest choice; other choices will be used in later chapters.

Let $A = A^x = C = I$ (II.15.5)

$B = -\lambda K$ (II.15.6)

$f = f(x), \; g = 0$ (II.15.7)

Equation (II.15.4) then becomes

$$L(\phi,\psi) = \langle\phi,\psi\rangle - \tfrac{1}{2}\langle\phi, \lambda K\phi\rangle - \tfrac{1}{2}\langle\psi,\psi\rangle + \langle\phi, f\rangle \qquad \text{(II.15.8)}$$

and the gradients are

$$\nabla_\phi L(\phi,\psi) = \psi - \lambda K\phi + f \qquad \text{(II.15.9)}$$

$$\nabla_\psi L(\phi,\psi) = \phi - \psi \qquad \text{(II.15.10)}$$

Setting $\nabla_\phi L = \nabla_\psi L = 0$ gives $(I - \lambda K)\phi = -f$, as required.

As $C = I$ is positive-definite, B is required to be positive-definite and symmetric to ensure that $L(\phi,\psi)$ is a strict convex-concave saddle functional; this will be considered after the dual extremum principles have been given. $\phi(x)$ and $f(x)$ must belong to the real space of square-integrable functions with inner product given by

$$\langle h_1, h_2\rangle = \int_R h_1(x)\, h_2(x)\, dx \qquad \text{(II.15.11)}$$

From the example at the end of section II.12, the dual extremum principles for this particular choice of operators are:

$$L(\phi_\beta, \psi_\beta) \le L(\phi_e, \psi_e) \le L(\phi_\alpha, \psi_\alpha) \qquad \text{(II.15.12)}$$

where

$$L(\phi_\beta, \psi_\beta) = \int_R \left\{\tfrac{1}{2}\phi_\beta(x)\, \lambda K\, \phi_\beta(x) - \tfrac{1}{2}(\psi_\beta(x))^2\right\} dx \qquad \text{(II.15.13)}$$

with $\psi_\beta - \lambda K\phi_\beta + f = 0$ (II.15.14)

$$L(\phi_\alpha, \psi_\alpha) = \int_R \left\{ -\tfrac{1}{2}\phi_\alpha(x)\, \lambda K\, \phi_\alpha(x) + \tfrac{1}{2}(\psi_\alpha(x))^2 + \phi_\alpha(x)\, f(x)\right\} dx \qquad \text{(II.15.15)}$$

with $\phi_\alpha - \psi_\alpha = 0$ (II.15.16)

$$L(\phi_e, \psi_e) = \int_R \tfrac{1}{2}\phi_e(x)\, f(x)\, dx \qquad \text{(II.15.17)}$$

with $(I - \lambda K)\phi_e(x) = -f(x)$ (II.15.18)

It is now necessary to find the conditions which ensure that B is symmetric and positive-definite.

From the example following definition (II.5.6), an integral operator K is self-adjoint, and therefore symmetric, if the kernel k(x,y) is symmetric; that is,

$$k(x,y) = k(y,x) \quad \forall \; x,y \in [a,b] \qquad \text{(II.15.19)}$$

B will be positive-definite if $\lambda < 0$ and $K > 0$ or $\lambda > 0$ and $K < 0$. From page 147 of (46), a symmetric integral operator K is positive-definite (negative-definite) if and only if

(a) All its eigenvalues are positive (negative) and

(b) The full orthonormal system of eigenfunctions is complete, where the eigenfunctions ϕn and eigenvalues μn are defined by the equation

$$K\phi_n = \mu_n \phi_n, \quad n \in \{1,2,\ldots\ldots\} \qquad \text{(II.15.20)}$$

From page 103 of (51), a system of functions $\{\phi_n(x)\}$, n = 1,2,... is complete if there does not exist a function with positive norm which is orthogonal to all of the functions of the system; two functions ϕ_1 and ϕ_2 are orthogonal on the domain $[a,b]$ if their inner product is zero (page 63 of (59)); that is,

$$\langle\phi_1, \phi_2\rangle = \int_a^b \phi_1(x)\, \phi_2(x) \; dx = 0 \qquad \text{(II.15.21)}$$

Finally, an orthonormal system of functions satisfies equation (II.15.21) for $\phi_1 \neq \phi_2$ and when $\phi_1 = \phi_2$,

$$\langle\phi_1, \phi_1\rangle = \int_a^b (\phi_1(x))^2 \; dx = 1 \qquad \text{(II.15.22)}$$

(page 75 of (46)).

CHAPTER II

The integral equations which will be used as illustrations in later chapters all arise from second order differential equations with given boundary values. From page 361 of (71), if the corresponding homogeneous differential equation with homogeneous boundary conditions has eigenfunctions $\{\varphi_n(x)\}$, n = 1,2,... and eigenvalues $\{\theta_n\}$, n = 1,2,..., then the eigenfunctions of the corresponding integral operator K are the same as those of the homogeneous differential equation, and the eigenvalues of K are the reciprocals of the eigenvalues θ_n.

Assuming that the system of eigenfunctions is a full, orthonormal, complete set, and the eigenvalues θ_n are either all positive or all negative, then K is a positive-definite or negative-definite operator. Bounds on K can then be found, as follows, noting that as the set of eigenfunctions is complete, the number of eigenvalues is infinite:

(i) Let all of the eigenvalues θ_n be positive; then

$0 < \theta_1 \leq \theta_2 \leq \theta_3 \leq \;\; ....$

The eigenvalues of K are then $\frac{1}{\theta_1}$, $\frac{1}{\theta_2}$, ... with

$$\frac{1}{\theta_1} \geq \frac{1}{\theta_2} \geq \frac{1}{\theta_3} \geq \;\; ...... \tag{II.15.23}$$

$$\text{As } \lim_{n\to\infty} \theta_n = \infty, \quad \lim_{n\to\infty} \frac{1}{\theta_n} = 0 \tag{II.15.24}$$

Hence the bounds in the integral operator K are

$$0 < K \leq \frac{1}{\theta_1} \tag{II.15.25}$$

(ii) Let all of the eigenvalues θ_n be negative; then

$$0 > \theta_1 \geq \theta_2 \geq \theta_3 \geq \;\; ...... \tag{II.15.26}$$

The eigenvalues of K are $\frac{1}{\theta_1}$, $\frac{1}{\theta_2}$, ... with

$$\frac{1}{\theta_1} \leq \frac{1}{\theta_2} \leq \frac{1}{\theta_3} \leq \;\; ...... \tag{II.15.27}$$

$$\text{As } \lim_{n \to \infty} \theta_n = -\infty, \quad \lim_{n \to \infty} \frac{1}{\theta_n} = 0 \tag{II.15.28}$$

Hence the bounds on the integral operator K are

$$\frac{1}{\theta_1} \leq K < 0 \tag{II.15.29}$$

The only integral operators which will be used in later chapters are ones which can be shown to be symmetric and positive-definite or negative-definite; the bounds on the operators will then be given by equation (II.15.25) or (II.15.29).

In later chapters it will be necessary to carry out conversions from integral equations to differential equations and vice-versa: the eigenvalues are found more easily from the differential equation, and the integral operators we are considering are bounded whereas a differential operator is unbounded. The rest of this section will therefore show how the conversions can be carried out and will also find the upper bounds for two particular integral operators.

<u>To convert an integral equation to its corresponding differential equation</u>

If we have an integral of the form $\int_{\phi(x)}^{\psi(x)} f(x,y)\,dy$, we can convert it to its corresponding differential form by using the following formula, from page 1098 of (40), twice:

$$\frac{d}{dx}\int_{\phi(x)}^{\psi(x)} f(x,y)\,dy = \left(\frac{d\psi}{dx}\right)\Big(f(\psi(x), x)\Big) - \left(\frac{d\phi}{dx}\right)\Big(f(\phi(x), x)\Big) + \int_{\phi(x)}^{\psi(x)} \frac{\partial f}{\partial x}\,dy \tag{II.15.30}$$

We can determine the boundary conditions associated with the differential equation by using the original integral equation and its derivative.

Example

Consider the integral equation

$$\phi(x) - \lambda\int_0^x y\,\phi(y)\,dy - \lambda\int_x^1 x\,\phi(y)\,dy$$

$$- \int_0^x y\,s(y)\,dy - \int_x^1 x\,s(y)\,dy - B_3\,x - A_3 = 0$$

where λ, B_3 and A_3 are real numbers (II.15.31)

Using equation (II.15.30),

$$\frac{d}{dx}\left[\begin{array}{l}\phi(x) - \lambda\int_0^x y\,\phi(y)\,dy - \lambda\int_x^1 x\,\phi(y)\,dy \\ - \int_0^x y\,s(y)\,dy + \int_x^1 x\,s(y)\,dy + B_3\,x + A_3\end{array}\right]$$

$$= \phi'(x) - \lambda(1.\,x\,\phi(x) - 0.x\,\phi(x) + \int_0^x 0\,dy)$$

$$- \lambda(0.x\,\phi(x) - 1.x\,\phi(x) + \int_x^1 \phi(y)\,dy)$$

$$- (1.x\,s(x) - 0.x\,s(x) + \int_0^x 0\,dy) - (0.x\,s(x) - 1.x\,s(x)$$

$$+ \int_x^1 s(y)\,dy) - B_3$$

$$= \phi'(x) - \lambda\int_x^1 \phi(y)\,dy - \int_x^1 s(y)\,dy - B_3$$

and the second derivative becomes

$$\frac{d}{dx}\left[\phi'(x) - \lambda\int_x^1 \phi(y)\,dy - \int_x^1 s(y)\,dy - B_3\right]$$

$$= \phi''(x) - \lambda(0.\,\phi(x) - 1.\phi(x)) - (0.s(x) - 1.s(x))$$

$$= \phi''(x) + \lambda\phi(x) + s(x)$$

Hence setting these two derivatives equal to zero, we have

$$\phi'(x) - \lambda\int_x^1 \phi(y)\,dy - \int_x^1 s(y)\,dy = B_3 \qquad \text{(II.15.32)}$$

and $\phi''(x) + \lambda\phi(x) = - s(x)$ (II.15.33)

CHAPTER II

<u>Example</u>

Consider the integral equation

$$\phi(x) - \lambda\int_0^x y\,\phi(y)\,dy - \lambda\int_x^1 x\,\phi(y)\,dy$$

$$- \int_0^x y\,s(y)\,dy - \int_x^1 x\,s(y)\,dy - B_3\,x - A_3 = 0$$

where λ, B_3 and A_3 are real numbers (II.15.31)

Using equation (II.15.30),

$$\frac{d}{dx}\left[\begin{array}{l}\phi(x) - \lambda\int_0^x y\,\phi(y)\,dy - \lambda\int_x^1 x\,\phi(y)\,dy \\ - \int_0^x y\,s(y)\,dy + \int_x^1 x\,s(y)\,dy + B_3\,x + A_3\end{array}\right]$$

$$= \phi'(x) - \lambda(1.\,x\,\phi(x) - 0.x\,\phi(x) + \int_0^x 0\,dy)$$

$$- \lambda(0.x\,\phi(x) - 1.x\,\phi(x) + \int_x^1 \phi(y)\,dy)$$

$$- (1.x\,s(x) - 0.x\,s(x) + \int_0^x 0\,dy) - (0.x\,s(x) - 1.x\,s(x)$$

$$+ \int_x^1 s(y)\,dy) - B_3$$

$$= \phi'(x) - \lambda\int_x^1 \phi(y)\,dy - \int_x^1 s(y)\,dy - B_3$$

and the second derivative becomes

$$\frac{d}{dx}\left[\phi'(x) - \lambda\int_x^1 \phi(y)\,dy - \int_x^1 s(y)\,dy - B_3\right]$$

$$= \phi''(x) - \lambda(0.\,\phi(x) - 1.\phi(x)) - (0.s(x) - 1.s(x))$$

$$= \phi''(x) + \lambda\phi(x) + s(x)$$

Hence setting these two derivatives equal to zero, we have

$$\phi'(x) - \lambda\int_x^1 \phi(y)\,dy - \int_x^1 s(y)\,dy = B_3 \qquad \text{(II.15.32)}$$

and $\phi''(x) + \lambda\,\phi(x) = -\,s(x)$ (II.15.33)

Equation (II.15.33) is the corresponding differential equation; we can find the associated boundary conditions by letting x = 0 and x = 1 in turn in equations (II.15.31) and (II.15.32), and choosing the equations which do not involve integrals. Putting x = 0 into equation (II.15.31) gives $\phi(0) = A_3$ whilst putting x = 1 into the same equation gives

$$\phi(1) - \lambda\int_0^1 y\phi(y)\,dy - \int_0^1 y\,s(y)\,dy - B_3 - A_3 = 0.$$

Putting x = 0 into equation (II.15.32) gives

$\phi'(0) - \lambda\int_0^1 \phi(y)\,dy - \int_0^1 s(y)\,dy = B_3$, whilst putting x = 1 into the same equation gives $\phi'(1) = B_3$. Hence choosing the equations which do not contain integrals gives the boundary conditions $\phi(0) = A_3$ and $\phi'(1) = B_3$.

<u>To convert a linear second order differential equation to its corresponding integral equation</u>

The non-homogeneous problem given by

$$\left(-\frac{d}{dx}\left(r(x)\frac{d}{dx}\right) + q(x) - \lambda\right)\phi(x) = s(x), \quad a \leq x \leq b$$

$$\begin{aligned} A_1\phi(a) + A_2\phi'(a) &= A_3 \\ B_1\phi(b) + B_2\phi'(b) &= B_3 \end{aligned} \tag{II.15.34}$$

$A_1, A_2, A_3, B_1, B_2, B_3 \in \mathbb{R}$

can be converted into the equivalent integral equation

$$\phi(x) - \lambda\int_a^b k(x,y)\,\phi(y)\,dy = \int_a^b k(x,y)\,s(y)\,dy + Ax + B \tag{II.15.35}$$

where k(x,y) is the Green's function for the problem and A and B are fixed by the boundary conditions in (II.15.34). The following method for finding k(x,y) is adapted from section 28 of (76) and section 1.3 of (30).

Let L be the general second order differential operator given by

$$L = -\frac{d}{dx}\left(r(x)\frac{d}{dx}\right) + q(x), \quad a \leq x \leq b \tag{II.15.36}$$

with the boundary conditions

$$A_1 \phi(a) + A_2 \phi'(a) = A_3$$
$$B_1 \phi(b) + B_2 \phi'(b) = B_3 \qquad \text{(II.15.37)}$$

$r(x)$ and $q(x)$ must be continuous on $[a,b]$: $r(x)$ must not vanish on $]a,b[$ and must have a continuous first derivative.

The Green's function $k(x,y)$ must satisfy the following conditions:

(a) $$-\frac{d}{dx}\left(r(x) \frac{dk(x,y)}{dx}\right) + q(x)\, k(x,y) = 0$$

(b) $$A_1 k^-\big|_{x=a} + A_2 \frac{dk^-}{dx}\Big|_{x=a} = B_1 k^+\big|_{x=b} + B_2 \frac{dk^+}{dx}\Big|_{x=b} = 0$$

(c) $$k^+\big|_{x=y} = k^-\big|_{x=y} \qquad \text{(II.15.38)}$$

(d) $$\frac{dk^+}{dx}\Big|_{x=y} - \frac{dk^-}{dx}\Big|_{x=y} = -\frac{1}{r(y)}$$

k^+ signifies that part of k for which $x \geq y$ and k^- signifies that part of k for which $x \leq y$. Then, for any $t(x)$ belonging to the domain of K,

$$K\, t(x) = \int_a^b k(x,y)\, t(y)\, dy$$
$$= \int_a^x k^+(x,y)\, t(y)\, dy + \int_x^b k^-(x,y)\, t(y)\, dy \qquad \text{(II.15.39)}$$

It should be noted that the Green's function exists if and only if the problem given by

$$\left(-\frac{d}{dx}\left(r(x) \frac{d}{dx}\right) + q(x)\right) \phi(x) = 0, \quad a \leq x \leq b$$
$$A_1 \phi(a) + A_2 \phi'(a) = 0 \qquad \text{(II.15.40)}$$
$$B_1 \phi(b) + B_2 \phi'(b) = 0$$

does not have a non-trivial solution.

The eigenfunctions and eigenvalues of L are found by solving the eigenvalue problem

$$\left(\left(- \frac{d}{dx} \quad r(x) \frac{d}{dx}\right) + q(x) - \theta\right) \psi = 0$$

$$A_1 \psi (a) + A_2 \; \psi'(a) = 0 \qquad \text{(II.15.41)}$$

$$B_1 \psi (b) + B_2 \; \psi'(b) = 0$$

then the eigenfunctions of K = eigenfunctions of $L = \psi_n$

and the eigenvalues of $K = \dfrac{1}{\text{eigenvalues of } L} = \dfrac{1}{\theta_n}$

Two examples which illustrate the conversions follow.

Example 1

Let $\phi(x)$ satisfy $- \phi''(x) - \lambda \phi(x) = s(x), \quad 0 \leq x \leq 1$

$$\phi(0) = A_3, \quad \phi'(1) = B_3 \qquad \text{(II.15.42)}$$

Comparing (II.15.42) with (II.15.36) and (II.15.37), we have

$r(x) = 1$, $q(x) = 0$, $A_1 = B_2 = 1$, $A_2 = B_1 = 0$ and $a = 0$, $b = 1$.

Using (II.15.38), $k(x,y)$ must satisfy

(a) $\quad - \dfrac{d^2k(x,y)}{dx^2} = 0$ $\qquad$ (b) $\quad k^- \Big|_{x = 0} = \dfrac{dk^+}{dx}\Big|_{x = 1} = 0$

(c) $\quad k^+\Big|_{x = y} = k^-\Big|_{x = y}$ $\qquad$ (d) $\quad \dfrac{dk^+}{dx}\Big|_{x = y} - \dfrac{dk^-}{dx}\Big|_{x = y} = -1$

From (a), we must have $k^-(x,y) = \alpha_1(y)\, x + \alpha_2(y)$ and

$k^+(x,y) = \beta_1(y)\, x + \beta_2(y)$.

From (b), $k^-\big|_{x = 0} = \alpha_2(y) = 0$; $\dfrac{dk^+}{dx}\Big|_{x = 1} = \beta_1(y) = 0$

Hence $k^-(x,y) = \alpha_1(y)\, x$, $\quad k^+(x,y) = \beta_2(y)$.

From (c), $y\, \alpha_1(y) = \beta_2(y)$

From (d), $0 - \alpha_1(y) = -1$

Hence $\alpha_1(y) = 1$ and $\beta_2(y) = y$

Therefore $k(x,y)$ is specified by

$$k(x,y) = \begin{cases} k^+(x,y) = y\,, & x \geq y \\ k^-(x,y) = x & x \leq y \end{cases} \qquad \text{(II.15.43)}$$

CHAPTER II

The integral equation corresponding to the differential equations given in (II.15.42) is therefore

$$\phi(x) - \lambda\int_0^x y\,\phi(y)\,dy - \lambda\int_x^1 x\,\phi(y)\,dy$$

$$= \int_0^x y\,s(y)\,dy + \int_x^1 x\,s(y)\,dy + Ax + B \qquad (II.15.44)$$

where A and B need to be determined.

Using equation (II.15.31) to differentiate (II.15.44), (as carried out earlier in this section),

$$\phi'(x) - \lambda\int_x^1 \phi(y)\,dy = \int_x^1 s(y)\,dy + A \qquad (II.15.45)$$

$\phi''(x) + \lambda\phi(x) = -\,s(x)$ as required.

From (II.15.44), $\phi(0) = B = A_3$

From (II.15.45), $\phi'(1) = A = B_3$

Hence the integral equation is

$$\phi(x) - \lambda\int_0^x y\,\phi(y)\,dy - \lambda\int_x^1 x\,\phi(y)\,dy$$

$$= \int_0^x y\,s(y)\,dy + \int_x^1 x\,s(y)\,dy + B_3x + A_3 \qquad (II.15.46)$$

To find the eigenfunctions ψ_n and eigenvalues Θ_n, we solve

$$-\frac{d^2\psi_n}{dx^2} - \Theta_n\psi_n = 0, \quad \psi_n(0) = \psi_n'(1) = 0 \qquad (II.15.47)$$

Then $\psi_n = \sin\frac{(2n-1)\pi}{2}x$, $n = 1,2,\ldots$ and $\Theta_n = \frac{(2n-1)^2\pi^2}{4}$

Hence the eigenvalues of K, as defined by equation (II.15.43) are

$$\mu_n = \frac{4}{(2n-1)^2\pi^2} \qquad n = 1,2,\ldots$$

The eigenfunctions form a full, orthonormal, complete system, and the eigenvalues are all positive, hence K is positive-definite and the bounds on K are

$$0 < K \leq \frac{4}{\pi^2} \qquad (II.15.48)$$

CHAPTER II

The integral equation corresponding to the differential equations given in (II.15.42) is therefore

$$\phi(x) - \lambda\int_0^x y\,\phi(y)\,dy - \lambda\int_x^1 x\,\phi(y)\,dy$$

$$= \int_0^x y\,s(y)\,dy + \int_x^1 x\,s(y)\,dy + Ax + B \qquad \text{(II.15.44)}$$

where A and B need to be determined.

Using equation (II.15.31) to differentiate (II.15.44), (as carried out earlier in this section),

$$\phi'(x) - \lambda\int_x^1 \phi(y)\,dy = \int_x^1 s(y)\,dy + A \qquad \text{(II.15.45)}$$

$\phi''(x) + \lambda\phi(x) = -\,s(x)$ as required.

From (II.15.44), $\phi(0) = B = A_3$

From (II.15.45), $\phi'(1) = A = B_3$

Hence the integral equation is

$$\phi(x) - \lambda\int_0^x y\,\phi(y)\,dy - \lambda\int_x^1 x\,\phi(y)\,dy$$

$$= \int_0^x y\,s(y)\,dy + \int_x^1 x\,s(y)\,dy + B_3x + A_3 \qquad \text{(II.15.46)}$$

To find the eigenfunctions ψ_n and eigenvalues Θ_n, we solve

$$-\frac{d^2\psi_n}{dx^2} - \Theta_n\psi_n = 0, \quad \psi_n(0) = \psi_n'(1) = 0 \qquad \text{(II.15.47)}$$

Then $\psi_n = \sin\frac{(2n-1)\pi}{2}x$, $n = 1,2,\ldots$ and $\Theta_n = \frac{(2n-1)^2\pi^2}{4}$

Hence the eigenvalues of K, as defined by equation (II.15.43) are

$$\mu_n = \frac{4}{(2n-1)^2\pi^2} \qquad n = 1,2,\ldots$$

The eigenfunctions form a full, orthonormal, complete system, and the eigenvalues are all positive, hence K is positive-definite and the bounds on K are

$$0 < K \leq \frac{4}{\pi^2} \qquad \text{(II.15.48)}$$

(or $0 \leq \int_0^1 \phi(x) \left(\int_0^1 k(x,y)\, \phi(y)\, dy\right) dx \leq \frac{4}{\pi^2} \int_0^1 (\phi(x))^2\, dx$)

<u>Example 2</u>

Let $\phi(x)$ satisfy $-\phi''(x) - \lambda\phi(x) = s(x), \quad a \leq x \leq b,$

$$\phi(a) = A_3, \quad \phi(b) = B_3 \tag{II.15.49}$$

Comparing (II.15.49) with (II.15.36) and (II.15.37), we have

$r(x) = 1,\ q(x) = 0,\ A_1 = B_1 = 1,\ A_2 = B_2 = 0.$

Using (II.15.38), k(x,y) must satisfy

(a) $-\frac{d^2k(x,y)}{dx^2} = 0$ (b) $k^-\big|_{x=a} = k^+\big|_{x=b} = 0$

(c) $k^+\big|_{x=y} = k^-\big|_{x=y}$ (d) $\frac{dk^+}{dx}\Big|_{x=y} - \frac{dk^-}{dx}\Big|_{x=y} = -1$

From (a): $k^-(x,y) = \alpha_1(y)\,x + \alpha_2(y),\quad k^+(x,y) = \beta_1(y)\,x + \beta_2(y)$

From (b): $\alpha_1(y)\,a + \alpha_2(y) = 0,\quad \beta_1(y)\,b + \beta_2(y) = 0$

hence $\alpha_2(y) = -\alpha_1(y)\,a$ and $\beta_2(y) = -\beta_1(y)\,b$

giving $k^-(x,y) = \alpha_1(y)\,(x - a),\quad k^+(x,y) = \beta_1(y)\,(x - b).$

From (c): $\alpha_1(y)\,(y - a) = \beta_1(y)\,(y - b)$

From (d): $\beta_1(y) - \alpha_1(y) = -1$

Solving (c) and (d) gives $\alpha_1(y) = \frac{b - y}{b - a} \qquad \beta_1(y) = \frac{a - y}{b - a}$

Therefore k(x,y) is specified by:

$$k(x,y) = \begin{cases} k^+(x,y) = \dfrac{(a - y)(x - b)}{(b - a)}, & x \geq y \\[2ex] k^-(x,y) = \dfrac{(b - y)(x - a)}{(b - a)}, & x \leq y \end{cases} \tag{II.15.50}$$

The integral equation corresponding to the differential equation given in (II.15.49) is then

$$\phi(x) - \lambda\int_a^x \frac{(a - y)(x - b)}{(b - a)}\,\phi(y)\,dy - \lambda\int_x^b \frac{(b - y)(x - a)}{(b - a)}\,\phi(y)\,dy$$

$$= \int_a^x \frac{(a-y)(x-b)}{(b-a)} s(y)\,dy + \int_x^b \frac{(b-y)(x-a)}{(b-a)} s(y)\,dy + Ax + B$$

where A and B are to be determined. (II.15.51)

From (II.15.48),

$\phi(a) = Aa + B = A_3$; $\phi(b) = Ab + B = B_3$

giving $A = \dfrac{B_3 - A_3}{b-a}$ and $B = \dfrac{bA_3 - aB_3}{b-a}$

Hence the integral equation is

$$\phi(x) - \lambda\int_a^x \frac{(a-y)(x-b)}{(b-a)} \phi(y)\,dy - \lambda\int_x^b \frac{(b-y)(x-a)}{(b-a)} \phi(y)\,dy$$

$$= \int_a^x \frac{(a-y)(x-b)}{(b-a)} s(y)\,dy + \int_x^b \frac{(b-y)(x-a)}{(b-a)} s(y)\,dy$$

$$+ \left(\frac{B_3 - A_3}{b-a}\right) x + \frac{bA_3 - aB_3}{b-a} \qquad \text{(II.15.52)}$$

A check by differentiating (II.15.52) twice gives (II.15.49), as required.

Solving the eigenfunction equation

$-\psi_n''(x) - \theta_n \psi_n(x) = 0$, $\psi_n(a) = \psi_n(b) = 0$ gives

$\psi_n = \sin \dfrac{n\pi x}{b-a}$ $n = 1,2,\ldots$ and $\theta_n = \left(\dfrac{n\pi}{b-a}\right)^2$, $n = 1,2,\ldots$

The eigenvalues of K, as defined by equation (II.15.47) are

$\mu_n = \left(\dfrac{b-a}{n\pi}\right)^2$, $n = 1,2,\ldots$ As the eigenfunctions form a full orthonormal complete set, K is positive-definite and so

$$0 < K \leq \left(\frac{b-a}{\pi}\right)^2 \qquad \text{(II.15.53)}$$

II.16 A simple convergence result for a bounded operator

Theorem II.16.1

Let P be a linear, self-adjoint operator, and let there exist real numbers q and Q such that

$$-\langle\phi,\phi\rangle < q\langle\phi,\phi\rangle \le \langle\phi, P\phi\rangle \le Q\langle\phi,\phi\rangle < \langle\phi,\phi\rangle \qquad \text{(II.16.1)}$$

If an iterative scheme defined by

$$u_{n+1} = Pu_n, \quad n = 0,1,2,3,\ldots, \quad u_n \neq 0 \qquad \text{(II.16.2)}$$

can be set up, then

$$\lim_{n\to\infty} \|u_{n+1}\| = 0 \qquad \text{(II.16.3)}$$

Proof

$$\|u_{n+1}\|^2 = \langle u_{n+1}, u_{n+1}\rangle$$

$$= \langle u_{n+1}, Pu_n\rangle \quad \text{from equation (II.16.2)}$$

$$\le \|u_{n+1}\|\,\|Pu_n\| \quad \text{by Lemma (II.4.1)}$$

then $\|u_{n+1}\| \le \|Pu_n\|$

$$\le \|P\|\,\|u_n\| \quad \text{by definition (II.5.10)}$$

$$= \max\{|q|, |Q|\}\,\|u_n\| \quad \text{by theorem (II.5.1)}$$

Let $R = \max\{|q|, |Q|\}$; $0 \le R < 1$ as $-1 < q \le Q < 1$

Then $\|u_{n+1}\| \le R\,\|u_n\|$

$$\le R^{n+1}\,\|u_0\|$$

$$\lim_{n\to\infty} \|u_{n+1}\| \le \lim_{n\to\infty} R^{n+1}\,\|u_0\|$$

$$= 0 \quad \text{as } 0 \le R < 1$$

This convergence result is a simplified version of convergence using the method of steepest descent, which is considered in the next section.

Note that equation (II.16.1) can be written

$$-\langle\phi,\phi\rangle < -Q\langle\phi,\phi\rangle \le \langle\phi, (-P)\phi\rangle \le -q\langle\phi,\phi\rangle < \langle\phi,\phi\rangle$$

and as $-1 < q \le Q < 1$ implies that $-1 < -Q \le -q < 1$, the theorem is valid for both positive and negative operators P.

CHAPTER II

II.17 Convergence of an iteration for bounded operators using the method of steepest descent

This section proves that, given an equation $P\phi_e = f$, where P is a linear, self-adjoint positive-definite operator, a particular iteration (the steepest descent iteration) will converge to ϕ_e. The proof depends on some lemmas, which are given first; the section is based on pages 224, 237-238 and 241 of (77).

Lemma II.17.1

Let H be a Hilbert space and P a positive-definite self-adjoint operator. Then the identity

$$(x,y)_p = \langle Px,y\rangle \qquad \text{(II.17.1)}$$

defines a new inner product on H.

Proof

We have to show that $(x,y)_p$ satisfies the inner product axioms given in definition (II.3.1):

(i) $(x_1, x_2)_p = \langle Px_1, x_2\rangle$

$= \langle Px_2, x_1\rangle$ as P is self-adjoint

$= (x_2, x_1)_p$

(ii) $(x_1, x_1)_p = \langle Px_1, x_1\rangle \geqslant 0$, with equality if and only if $x_1 = 0$ as P is positive-definite.

(iii) $(x_1, \alpha x_2 + \beta x_3)_p = \langle Px_1, \alpha x_2 + \beta x_3\rangle$

$= \alpha\langle Px_1, x_2\rangle + \beta\langle Px_1, x_3\rangle$

$= \alpha(x_1, x_2)_p + \beta(x_1, x_3)_p$

$\forall\ x_1, x_2, x_3 \in H$ and $\forall\ \alpha, \beta \in \mathbb{R}$

Hence $(x,y)_p$ satisfies the inner product axioms.

Lemma II.17.2

If H is a Hilbert space and P is a positive-definite self-adjoint operator on H, then

$$\langle Px,y\rangle^2 \leq \langle Px,x\rangle\langle Py,y\rangle \tag{II.17.2}$$

Proof

$\langle Px,y\rangle^2 = \{(x,y)\ p\}^2$ from (II.17.1)

$\leq (x,x)p\ (y,y)p$ by Lemma (II.4.1)

$= \langle Px,x\rangle\langle Py,y\rangle$ by equation (II.17.1)

Hence $\langle Px,y\rangle^2 \leq \langle Px,x\rangle\langle Py,y\rangle$

Lemma II.17.3

If P is a self-adjoint positive-definite operator defined on a Hilbert space H then

$$\| P(\phi_{2n} - \phi_e) \|^4 \leq \langle \phi_{2n} - \phi_e, P(\phi_{2n} - \phi_e)\rangle\langle P(\phi_{2n} - \phi_e), PP(\phi_{2n} - \phi_e)\rangle \tag{II.17.3}$$

Proof

This follows by letting $x = \phi_{2n} - \phi_e$, $y = P(\phi_{2n} - \phi_e)$ in equation (II.17.2) and noting that $\langle \phi, \phi\rangle^2 = \|\phi\|^4$ by definition (II.4.3).

Lemma II.17.4

$(I + \varepsilon_{2n}P)$ is a self-adjoint, positive-definite operator whenever P is a positive-definite self-adjoint operator, where $\varepsilon_{2n} = - \dfrac{\langle Z_{2n}, Z_{2n}\rangle}{\langle Z_{2n}, PZ_{2n}\rangle}$ (II.17.4)

and $Z_{2n} = P(\phi_{2n} - \phi_e)$ (II.17.5)

Proof

(i) P is symmetric $\Rightarrow \langle \phi_1, P\phi_2\rangle = \langle \phi_2, P\phi_1\rangle \ \forall \phi_1, \phi_2 \in H$ (II.17.6)

$\langle \phi_1, (I + \varepsilon_{2n}P)\phi_2\rangle = \langle \phi_1, \phi_2\rangle + \varepsilon_{2n}\langle \phi_1, P\phi_2\rangle$

$= \langle \phi_2, \phi_1\rangle + \varepsilon_{2n}\langle \phi_2, P\phi_1\rangle$ from (II.17.6)

$= \langle \phi_2, (I + \varepsilon_{2n}P)\phi_1\rangle$

Hence $I + \varepsilon_{2n}P$ is symmetric.

(ii) As P is self-adjoint there exist real numbers q and Q such that

$$q\langle\phi,\phi\rangle \le \langle\phi, P\phi\rangle \le Q\langle\phi,\phi\rangle \quad \text{(II.17.7)}$$

As ε_{2n} is a real number, equation (II.17.7) implies that

$$(1 + \varepsilon_{2n}q)\langle\phi,\phi\rangle \le \langle\phi,(I + \varepsilon_{2n}P)\phi\rangle \le (1 + \varepsilon_{2n}Q)\langle\phi,\phi\rangle$$

hence $(I + \varepsilon_{2n}P)$ is bounded and is therefore self-adjoint, using definition (II.5.6).

(iii) P is positive definite $\Rightarrow \langle\phi, P\phi\rangle \ge 0 \quad \forall \ \phi \in H$, (II.17.8)

with equality if and only if $\phi = 0$.

$(I + \varepsilon_{2n}P)$ is positive definite if

$$\langle\phi_{2n} - \phi_e, (I + \varepsilon_{2n}P)(\phi_{2n} - \phi_e)\rangle \ge 0 \ \forall \ (\phi_{2n} - \phi_e) \in H \quad \text{(II.17.9)}$$

with equality if and only if $\phi_{2n} = \phi_e$.

Now $\langle\phi_{2n} - \phi_e, (I + \varepsilon_{2n}P)(\phi_{2n} - \phi_e)\rangle$

$$= \langle\phi_{2n} - \phi_e, \phi_{2n} - \phi_e\rangle + \varepsilon_{2n}\langle\phi_{2n} - \phi_e, P(\phi_{2n} - \phi_e)\rangle$$

$$= \|\phi_{2n} - \phi_e\|^2 - \frac{\langle z_{2n}, z_{2n}\rangle\langle\phi_{2n} - \phi_e, z_{2n}\rangle}{\langle z_{2n}, P z_{2n}\rangle} \quad \text{(II.17.10)}$$

using (II.17.4) and (II.17.5)

By (II.17.3) and (II.17.5)

$$\|z_{2n}\|^4 \le \langle\phi_{2n} - \phi_e, z_{2n}\rangle\langle z_{2n}, P z_{2n}\rangle$$

hence $\dfrac{\langle z_{2n}, z_{2n}\rangle}{\langle z_{2n}, Pz_{2n}\rangle} \le \dfrac{\langle\phi_{2n} - \phi_e, z_{2n}\rangle}{\langle z_{2n}, z_{2n}\rangle}$

and so $\dfrac{\langle z_{2n}, z_{2n}\rangle\langle\phi_{2n} - \phi_e, z_{2n}\rangle}{\langle z_{2n}, P z_{2n}\rangle} \le \dfrac{\langle\phi_{2n} - \phi_e, z_{2n}\rangle^2}{\langle z_{2n}, z_{2n}\rangle}$

$$\le \frac{\|\phi_{2n} - \phi_e\|^2 \ \|z_{2n}\|^2}{\|z_{2n}\|^2}$$

using lemma (II.4.1)

Therefore, $\dfrac{\langle z_{2n}, z_{2n}\rangle\langle\phi_{2n} - \phi_e, z_{2n}\rangle}{\langle z_{2n}, P z_{2n}\rangle} \le \|\phi_{2n} - \phi_e\|^2$

or $\|\phi_{2n} - \phi_e\|^2 - \dfrac{\langle z_{2n}, z_{2n}\rangle\langle\phi_{2n} - \phi_e, z_{2n}\rangle}{\langle z_{2n}, P z_{2n}\rangle} \ge 0$

or $\langle\phi_{2n} - \phi_e, (I + \varepsilon_{2n}P)(\phi_{2n} - \phi_e)\rangle \ge 0$ (II.17.11)

As $\langle \phi, P\phi \rangle = 0$ if and only if $\phi = 0$,

$\langle \phi_{2n} - \phi_e, P(\phi_{2n} - \phi_e) \rangle = 0$ if and only if $\phi_{2n} = \phi_e$

and so equality in equation (II.17.11) only occurs if $\phi_{2n} = \phi_e$.

Hence $(I + \varepsilon_{2n}P)$ is positive definite.

Lemma II.17.5

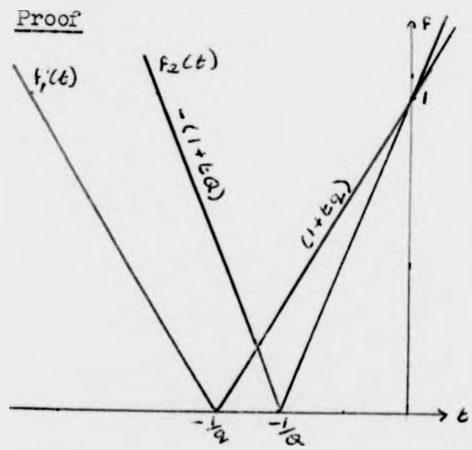
If t, q and Q are real numbers, $0 < q < Q$ then

$\underset{\text{with respect to } t}{\text{Minimum}} \left(\text{Max} \left\{ |1 + tq|, |1 + tQ| \right\} \right)$ occurs when

$$t = \frac{-2}{q + Q} \qquad \text{(II.17.12)}$$

giving $$|1 + tq| = |1 + tQ| = \frac{Q - q}{Q + q} \qquad \text{(II.17.13)}$$

Proof



Let $f_1(t) = |1 + tq|$
$= 1 + tq, \; t \geq -\frac{1}{q}$
$= -(1 + tq), \; t \leq -\frac{1}{q}$

Let $f_2(t) = |1 + tQ|$
$= 1 + tQ, \; t \geq -\frac{1}{Q}$
$= -(1 + tQ), \; t \leq -\frac{1}{Q}$

Figure II.17.1

From figure (II.17.1), it is clear that the maximum of $\{|1 + tq|, |1 + tQ|\}$ occurs on the red line, and that the minimum of this maximum occurs when

$$-(1 + tQ) = (1 + tq) \qquad \text{(II.17.14)}$$

Solving (II.17.14) for t gives

$$t = \frac{-2}{q + Q}$$

Then $\left|1 + tq\right| = \left|1 - \frac{2q}{q + Q}\right| = \left|\frac{Q - q}{Q + q}\right| = \frac{Q - q}{Q + q}$ as $0 < q \leq Q$

and $\left|1 + tQ\right| = \left|1 - \frac{2Q}{q + Q}\right| = \left|\frac{q - Q}{Q + q}\right| = \frac{Q - q}{Q + q}$ as $0 < q \leq Q$

Theorem II.17.1

Let P be a linear, self-adjoint, positive definite operator on a Hilbert space H. Then the steepest descent iteration, specified by

$$\phi_{2n+2} = \phi_{2n} + \varepsilon_{2n} \, Z_{2n} \tag{II.17.15}$$

$$\varepsilon_{2n} = -\frac{\langle Z_{2n}, Z_{2n}\rangle}{\langle Z_{2n}, PZ_{2n}\rangle} \tag{II.17.16}$$

$$Z_{2n} = P\phi_{2n} - f \tag{II.17.17}$$

converges to the unique solution ϕ_e of

$$P\phi_e = f \tag{II.17.18}$$

Proof

The theorem is proved by showing that

$$\lim_{n \to \infty} \|\phi_{2n+2} - \phi_e\| = 0 \tag{II.17.19}$$

As P is a self-adjoint, positive-definite operator, there exists positive real numbers q and Q such that

$$0 < q\langle\phi, \phi\rangle \leq \langle\phi, P\phi\rangle \leq Q\langle\phi, \phi\rangle \tag{II.17.20}$$

$$\begin{aligned}\phi_{2n+2} - \phi_e &= \phi_{2n} - \phi_e + \varepsilon_{2n} \, Z_{2n} \text{ from (II.17.15)} \\ &= \phi_{2n} - \phi_e + \varepsilon_{2n} \, (P(\phi_{2n} - \phi_e)) \text{ from (II.17.17), (II.17.18)} \\ &= (I + \varepsilon_{2n}P)(\phi_{2n} - \phi_e)\end{aligned} \tag{II.17.21}$$

By lemma (II.17.4), $(I + \varepsilon_{2n}P)$ is a self-adjoint, positive-definite operator and is therefore bounded; the bounds are given by

$$\begin{aligned}0 < (1 + \varepsilon_{2n}\, q)\langle\phi, \phi\rangle &\leq \langle\phi, (I + \varepsilon_{2n}\, P)\phi\rangle \\ &\leq (1 + \varepsilon_{2n}\, Q)\langle\phi, \phi\rangle\end{aligned} \tag{II.17.22}$$

$$\|\phi_{2n+2} - \phi_e\| = \|(I + \varepsilon_{2n} P)(\phi_{2n} - \phi_e)\|$$
$$= \|(I + \varepsilon_{2n} P)^{n+1} (\phi_o - \phi_e)\|$$
$$\leq \|I + \varepsilon_{2n} P\|^{n+1} \|\phi_o - \phi_e\| \qquad \text{(II.17.23)}$$

using definition (II.4.12)

By theorem (II.5.1),

$$\|I + \varepsilon_{2n} P\| = \max \{|1 + \varepsilon_{2n} q|, |1 + \varepsilon_{2n} Q|\} \qquad \text{(II.17.24)}$$

By lemma (II.17.5), this bound will be a minimum if

$$\varepsilon_{2n} = \frac{-2}{Q + q}, \text{ and } |1 + \varepsilon_{2n} q| = |1 + \varepsilon_{2n} Q| = \frac{Q - q}{Q + q}$$

Then $\|I + \varepsilon_{2n} P\| = \dfrac{Q - q}{Q + q}$ (II.17.25)

$$\lim_{n\to\infty} \|\phi_{2n+2} - \phi_e\| \leq \lim_{n\to\infty} \|I + \varepsilon_{2n} P\|^{n+1} \|\phi_o - \phi_e\|$$
$$= \lim_{n\to\infty} \left(\frac{Q - q}{Q + q}\right)^{n+1} \|\phi_o - \phi_e\|$$
$$= 0 \text{ as } 0 \leq \frac{Q - q}{Q + q} < 1$$

CHAPTER II

II.18 Convergence of an iteration for unbounded operators using the method of steepest descent

This section proves a convergence result analogous to the result in the previous section, but this time for unbounded operators. The proof depends on several lemmas and definitions, which are given first. The section is based on pages 88-89, 91, 99, 212, 214 and 252-255 of (77).

Definition II.18.1

If two operators S and T satisfy $D(S) \subset D(T)$ and $Sx = Tx \ \forall \ x \in D(S)$ then T is an extension of S, and $S \subset T$.

Definition II.18.2

Let T be a linear operator with a dense domain D(T) (see definition (II.5.12)). If $\langle Tx,y \rangle = \langle x, Ty \rangle \ \forall \ x,y \in D(T)$ then T is a symmetric operator. If D(T) is a Hilbert space, T is a self-adjoint operator.

Lemma II.18.1

Let Q be a densely defined symmetric operator with domain $D(Q) \subset H(Q)$ where H(Q) is a Hilbert space. Then any extension of Q whose domain is H(Q) is self-adjoint.

Proof: This follows from definition (II.18.2)

Definition II.18.3

A mapping T is one-one if, $\forall \ x_1, x_2 \in D(T)$,

$$Tx_1 = Tx_2 \Rightarrow x_1 = x_2 \qquad \text{(II.18.1)}$$

In this case there exists the inverse mapping

$T^{-1} : R(T) \rightarrow D(T)$

where R(T) signifies the image set of T in the domain D(T): and

$$T^{-1} Tx = x \ \forall \ x \in D(T), \quad TT^{-1} y = y \ \forall \ y \in R(T) \qquad \text{(II.18.2)}$$

Lemma II.18.2

Let $Q : D(Q) \to R(Q)$ be a linear operator with inverse Q^{-1}. Then Q^{-1} is also linear.

Proof

Let $x_1, x_2 \in D(Q)$ and $y_1, y_2 \in R(Q)$, such that

$$Qx_1 = y_1 \quad \text{and} \quad Qx_2 = y_2 \tag{II.18.3}$$

$$\text{Then } x_1 = Q^{-1}y_1 \quad \text{and} \quad x_2 = Q^{-1}y_2 \tag{II.18.4}$$

As Q is linear, by definition (II.5.2), there exist real numbers α, β such that

$$Q(\alpha x_1 + \beta x_2) = \alpha Qx_1 + \beta Qx_2$$

$$= \alpha y_1 + \beta y_2 \text{ from (II.18.3)}$$

$$\text{Then } Q^{-1}(\alpha y_1 + \beta y_2) = Q^{-1} Q(\alpha x_1 + \beta x_2)$$

$$= \alpha x_1 + \beta x_2$$

$$= \alpha Q^{-1}y_1 + \beta Q^{-1}y_2 \text{ from (II.18.4)}$$

Hence Q^{-1} is linear.

Lemma II.18.3

Let Q be a linear operator. Then the inverse Q^{-1} exists if and only if

$$Qx = 0 \Rightarrow x = 0 \tag{II.18.5}$$

Proof

(i) Assume Q^{-1} exists, and suppose $Qx = 0$

Then $Q^{-1} Qx = Q^{-1}(0)$

$$\text{Now } Q^{-1} Qx = x, \text{ by equation (II.18.2)} \tag{II.18.6}$$

Also, from Lemma (II.18.2), Q^{-1} is linear, hence

$$Q^{-1}(\alpha x) = \alpha Q^{-1}(x) \tag{II.18.7}$$

Setting $\alpha = 0$ in equation (II.18.7) gives

$$Q^{-1}(0) = 0 \tag{II.18.8}$$

Thus $Q^{-1} Qx = Q^{-1}(0) \Leftrightarrow x = 0$

Hence if Q^{-1} exists, $Qx = 0 \Rightarrow x = 0$.

(ii) Assume $Qx = 0 \Rightarrow x = 0$ (II.18.9)

Let $Qx_1 = Qx_2$; since Q is linear,

$Q(x_1 - x_2) = Qx_1 - Qx_2 = 0$, so

$x_1 - x_2 = 0$, by equation (II.18.9) (II.18.10)

Therefore $Qx_1 = Qx_2 \Rightarrow x_1 = x_2$.

Q is then one-one and by definition (II.18.3), Q^{-1} exists).

Lemma II.18.4

Let Q be a linear operator, $Q : D(Q) \rightarrow R(Q)$, and let there exist a real number Y such that

$$0 < Y^2 \langle x,x \rangle \leq \langle x, Qx \rangle \qquad \text{(II.18.11)}$$

Then Q^{-1} exists and $\| Q^{-1} x \| \leq \frac{1}{Y^2} \| x \|$ (II.18.12)

Proof

By Lemma (II.4.1), as Q is positive-definite,

$$\langle x, Qx \rangle^2 \leq \langle Qx, Qx \rangle \langle x,x \rangle$$
$$= \| Qx \|^2 \| x \|^2$$

then $0 < Y^2 \| x \|^2 \leq \langle x, Qx \rangle \leq \| Qx \| \| x \|$

or $\| Qx \| \geq Y^2 \| x \|$ (II.18.13)

Using equation (II.18.13), $Qx = 0 \Rightarrow x = 0$. By Lemma (II.18.3), Q^{-1} exists, such that $Q^{-1} : R(Q) \rightarrow D(Q)$.

For any $y \in R(Q)$, let $x = Q^{-1}y$ (II.18.14)

Then $\| Qx \| = \| QQ^{-1}y \| = \| y \| \geq Y^2 \| x \|$
$= Y^2 \| Q^{-1}y \|$

Hence $\| Q^{-1}y \| \leq \frac{1}{Y^2} \| y \|$

or $\| Q^{-1}x \| \leq \frac{1}{Y^2} \| x \|$ as required.

Lemma II.18.5

Let $G^{-1} \in H(Q)$ be an extension of Q where Q is defined as in Lemmas (II.18.1) and (II.18.4). Then G exists, satisfies $\| G \| \leq \frac{1}{\gamma^2}$ (II.18.15)

and is self-adjoint.

Proof

This follows from Lemmas (II.18.1) and (II.18.4) by noting that $Q = G^{-1}$ on $D(Q)$, and G is self-adjoint as it is bounded.

Lemma II.18.6

If u,v belong to an inner product space X then

$$\| u \|^2 = \sup_{v \neq 0} \frac{|\langle u,v \rangle|^2}{\langle v,v \rangle} \tag{II.18.16}$$

Proof

By Lemma (II.4.1) and definition (II.4.3),

$$\frac{|\langle u,v \rangle|^2}{\langle v,v \rangle} \leq \frac{\| u \|^2 \| v \|^2}{\| v \|^2} = \| u \|^2 \tag{II.18.17}$$

Therefore $\sup_{v \neq 0} \frac{|\langle u,v \rangle|^2}{\langle v,v \rangle} \leq \| u \|^2$ (II.18.18)

Suppose $\sup_{v \neq 0} \frac{|\langle u,v \rangle|^2}{\langle v,v \rangle} < \| u \|^2$,

that is, there exists a real number $K \in \,]0,1[$

such that $\sup_{v \neq 0} \frac{|\langle u,v \rangle|^2}{\langle v,v \rangle} = K \| u \|^2$ (II.18.19)

Let $V = \varepsilon u$, $\varepsilon \in \mathbb{R}$, then

$$\frac{|\langle u,v \rangle|^2}{\langle v,v \rangle} = \frac{|\langle u, \varepsilon u \rangle|^2}{\langle \varepsilon u, \varepsilon u \rangle} = \| u \|^2$$

which contradicts assumption (II.18.19).

Hence $\| u \|^2 = \sup_{v \neq 0} \frac{|\langle u,v \rangle|^2}{\langle v,v \rangle}$

CHAPTER II

Theorem II.18.1

Let P and Q be two closed symmetric operators with dense domains on a Hilbert space H such that $D(Q) \subseteq D(P)$. Assume that P and Q satisfy, for $m, M, Y \in \mathbb{R}$,

(a) $\langle x, Qx\rangle \geq Y^2 \langle x,x\rangle \quad \forall \; x \in D(Q)$ (II.18.20)

(b) $m \langle Qx,x\rangle \leq \langle Px,x\rangle \leq M \langle Qx,x\rangle \; \forall \; x \in D(Q)$ (II.18.21)

Then the iteration $\phi_{2n+2} = \phi_{2n} + \varepsilon_{2n} \, Z_{2n}$ (II.18.22)

$$Q \, Z_{2n} = P\phi_{2n} - f \qquad \text{(II.18.23)}$$

$$\varepsilon_{2n} = - \frac{\langle P\phi_{2n} - f, \; Z_{2n}\rangle}{\langle P \, Z_{2n}, \; Z_{2n}\rangle} \qquad \text{(II.18.24)}$$

will converge to the unique solution ϕ_e of

$$P\phi_e = f \qquad \text{(II.18.25)}$$

Proof

From (II.18.20), (II.18.21),

$\langle Px,x\rangle \geq m Y^2 \langle x,x\rangle, \quad x \in D(Q)$ (II.18.26)

We introduce the auxiliary Hilbert space $H(Q)$,

$D(Q) \subseteq H(Q) \subseteq H$ (II.18.27)

and the inner product

$(u,v)_Q = \langle Qu,v\rangle, \quad u,v \in D(Q)$ (II.18.28)

which was shown to be an inner product in Lemma (II.17.1).

The norm is denoted by $[\![\;]\!]$, where

$[\![u]\!]^2 = (u,u)_Q$ (II.18.29)

Let there be two self-adjoint extensions of Q, $\hat{Q}$ and G^{-1} (see Lemma (II.18.1)), then G exists and satisfies $\|G\| \leq \frac{1}{Y^2}$, and is self-adjoint (see Lemma (II.18.5)).

By the definition of an extension, definition (II.18.1),

$Q = \hat{Q} = G^{-1}$ on $D(Q)$ (II.18.30)

and therefore $QG = GQ = \hat{Q}G = G\hat{Q} = I$ on $H(Q)$ (II.18.31)

Also,

$$m(x,x)_Q = m\langle \hat{Q}\, x,x\rangle \leq \langle Px,x\rangle \leq M\langle \hat{Q}\, x,x\rangle = M(x,x)_Q, \quad x \in D(Q) \tag{II.18.32}$$

(from (II.18.21), (II.18.28) and (II.18.30))

and

$$\langle Px,y\rangle = \langle \hat{Q}G\, Px,y\rangle, \quad x,y \in D(Q) \tag{II.18.33}$$

(from (II.18.31)).

If we consider the space $H(\hat{Q})$ constructed from $\langle \hat{Q}x,x\rangle$, we have

$D(Q) \subseteq H(Q) \subsetneq H$,

$D(Q) \subseteq D(\hat{Q}) \subseteq H(\hat{Q})$, while

$$\langle \hat{Q}x,x\rangle = \| x \|^2_{\hat{Q}} = \| x \|^2_Q = \langle Qx,x\rangle \tag{II.18.34}$$

for all $x \in D(Q)$ (from (II.18.30)) and $D(Q)$ is dense in $H(Q)$, $H(\hat{Q})$ and H. Then convergence in the $\|\cdot\|_Q$ and $\|\ \|_{\hat{Q}}$ norms are equivalent so $H(Q) = H(\hat{Q})$.

From equations (II.18.28), (II.18.30) and (II.18.33),

$$\langle Px,x\rangle = \langle Q\, G\, Px,x\rangle = (G\, Px,x)_Q, \quad x \in D(Q) \tag{II.18.35}$$

Then, using equation (II.18.32),

$$m(x,x)_Q \leq (G\, Px,x)_Q \leq M(x,x)_Q \tag{II.18.36}$$

Using Lemma (II.18.6), with $X = D(Q)$, $u = G\, Px$, $v = y$, and the inner product and norm given by equations (II.18.28) and (II.18.29),

$$\begin{aligned} \| G\, Px \|^2 &= \sup_{y\neq 0} \frac{|(G\, Px,\, y)_Q|^2}{(y,y)_Q} \\ &= \sup_{y\neq 0} \frac{|\langle Q\, G\, Px,y\rangle|^2}{\langle Qy,y\rangle} \quad \text{using equation (II.18.28)} \\ &= \sup_{y\neq 0} \frac{|\langle Px,y\rangle|^2}{\langle Qy,y\rangle} \quad (\text{as } QG = I) \end{aligned} \tag{II.18.37}$$

As P is a positive-definite operator, from Lemma (II.17.2) we have

$$|\langle Px,y\rangle|^2 \leq \langle Px,x\rangle\langle Py,y\rangle \tag{II.18.38}$$

Hence, for $x \in D(Q)$,

$$\| GPx \|^2 \leq \sup_{y \neq 0} \frac{\langle Px,x \rangle \langle Py,y \rangle}{\langle Qy,y \rangle} \tag{II.18.39}$$

From equation (II.18.21), $\dfrac{\langle Py,y \rangle}{\langle Qy,y \rangle} \leq M$ (II.18.40)

Therefore

$$\begin{aligned} \| G\,Px \|^2 &\leq \sup_{y \neq 0} M \langle Px,x \rangle \\ &= M \langle Px,x \rangle \\ &\leq M^2 (x,x)_Q \quad \text{from equation (II.18.32)} \\ &= M^2 \| x \|^2 \end{aligned}$$

Hence GP is a bounded and uniformly continuous operator on $D(Q)$; it has a unique bounded extension with the same bound on $H(Q)$.

$$\begin{aligned} \text{Now } (G\,Px,y)_Q &= \langle Q\,G\,Px,y \rangle \\ &= \langle Px,y \rangle \\ &= \langle x, Py \rangle \quad \text{as P is symmetric} \\ &= \langle G\,Qx, Py \rangle \quad \text{as } GQ = I \text{ on } H(Q) \\ &= \langle Qx, G\,Py \rangle \quad \text{as G is self-adjoint} \\ &= (x, G\,Py)_Q \end{aligned}$$

Thus GP is self-adjoint as it is symmetric and bounded. Therefore the extension of GP, which we write as $\widehat{GP}$, has the property that the quadratic form is always real, and hence is a bounded self-adjoint operator on $H(Q)$.

To $\widehat{GP}$ we can apply the method of steepest descent for bounded operators, that is theorem (II.17.1). By this theorem, the iteration given by the equations

$$\phi_{2n+2} = \phi_{2n} + \varepsilon_{2n} z_{2n} \tag{II.18.41}$$

$$z_{2n} = \widehat{GP}\, \phi_{2n} - Gf \tag{II.18.42}$$

$$\varepsilon_{2n} = - \frac{(z_{2n}, z_{2n})_Q}{(\widehat{GP}\, z_{2n}, z_{2n})_Q} \tag{II.18.43}$$

converges to the solution of $\widehat{GP}\, \phi_e = Gf$ (II.18.44)

Suppose now that in some set $K \subseteq D(P)$, $x \in K$ implies $Px \in R(Q)$ and $G\,Px \in K$, and $Gh \in K$.

Then, if $\phi_{2n} \in K$, $GP\phi_{2n} \in K$ and $Gf \in K$; hence $Z_{2n} \in K$.

Therefore, equation (II.18.42) can be written

$$\begin{aligned} Q\,Z_{2n} &= Q\,\hat{GP}\,\phi_{2n} - Q\,G\,f \\ &= Q\,GP\,\phi_{2n} - Q\,G\,f \\ &= P\phi_{2n} - f \end{aligned} \qquad \text{(II.18.45)}$$

Equation (II.18.43) is equivalent to

$$\begin{aligned} \varepsilon_{2n} &= -\frac{\langle Q\,Z_{2n},\ Z_{2n}\rangle}{\langle Q\,\hat{GP}\,Z_{2n},\ Z_{2n}\rangle} \\ &= -\frac{\langle P\phi_{2n} - f,\ Z_{2n}\rangle}{\langle P\,Z_{2n},\ Z_{2n}\rangle} \end{aligned} \qquad \text{(II.18.46)}$$

and equation (II.18.44) becomes

$$\begin{aligned} Q\,\hat{GP}\,\phi_e &= Q\,G\,f \\ \text{or} \quad P\phi_e &= f \end{aligned} \qquad \text{(II.18.47)}$$

thus giving the iteration

$$\begin{aligned} \phi_{2n+2} &= \phi_{2n} + \varepsilon_{2n}\ Z_{2n} \\ QZ_{2n} &= P\phi_{2n} - f \\ \varepsilon_{2n} &= -\frac{\langle P\phi_{2n} - f,\ Z_{2n}\rangle}{\langle P\,Z_{2n},\ Z_{2n}\rangle} \end{aligned}$$

converges to the unique solution of $P\phi_e = f$, as required.

CHAPTER II

II.19 Cobweb Iteration

In the next chapter, a new iterative scheme is developed by which successively better approximations to the stationary value of a functional $L(\phi,\psi)$ can be found. Our aim is to use this analogy for quadratic functionals and so the convergence of cobweb iteration when applied to real linear functions is considered here.

It is assumed that there exist two linear non-parallel lines in the xy plane, defined by

$$y = mx + b \quad \text{and} \quad y = nx + C, \qquad n < 0 < m \tag{II.19.1}$$

which intersect at (x_e, y_e). The sketches in Figure (II.19.1) indicate that if $\left|\frac{m}{n}\right| < 1$, for convergence it is necessary to iterate in an anti-clockwise direction, and if $\left|\frac{m}{n}\right| > 1$, it is necessary to iterate in a clockwise direction; if $|m| = |n|$, no convergence is possible. The purpose of this section is to show that the criteria suggested by the sketches in figure (II.19.1) are correct. It should also be obvious from the sketches that convergence does not depend on the point at which the iterations start.

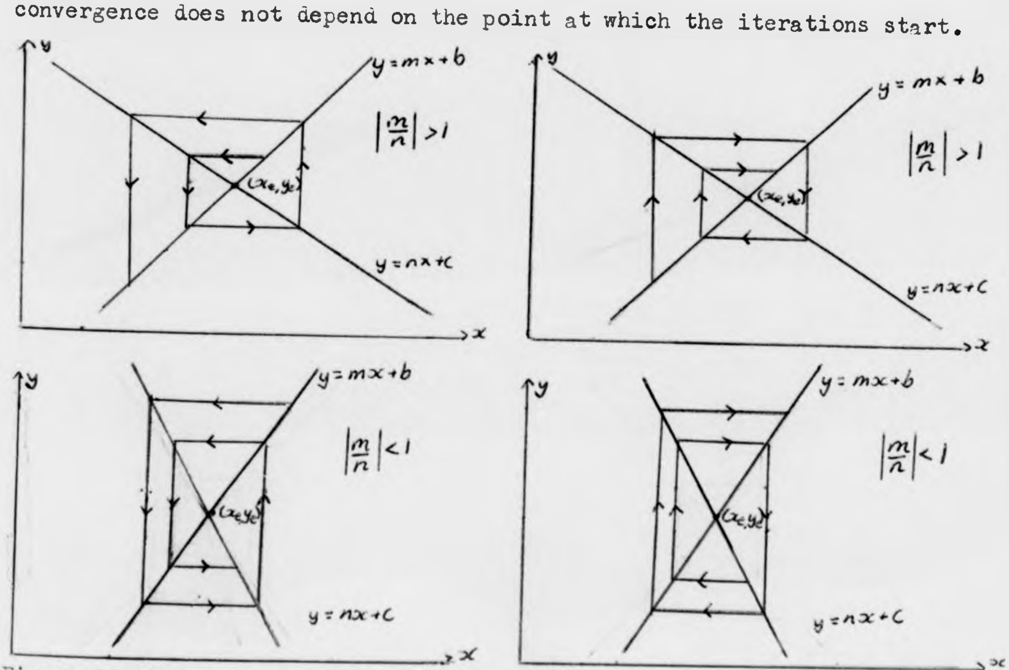
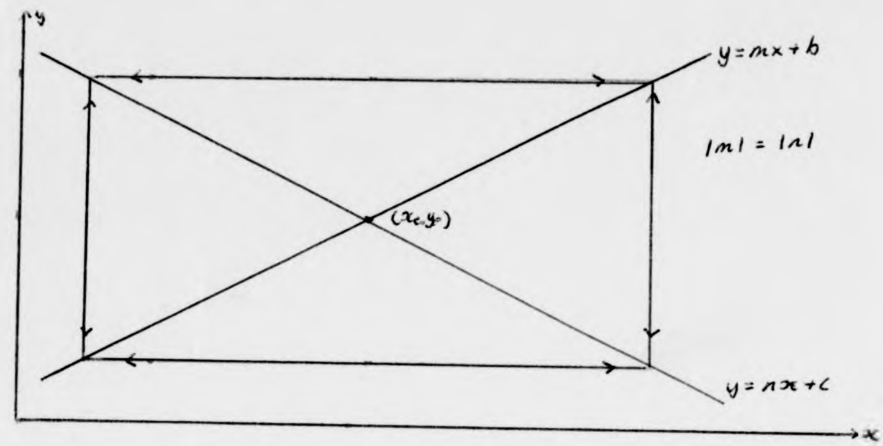


Figure II.19.1



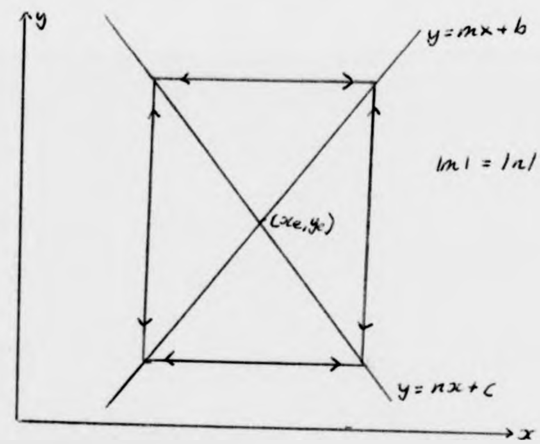


Figure II.19.1 (cont'd

(i) Anticlockwise Iterations

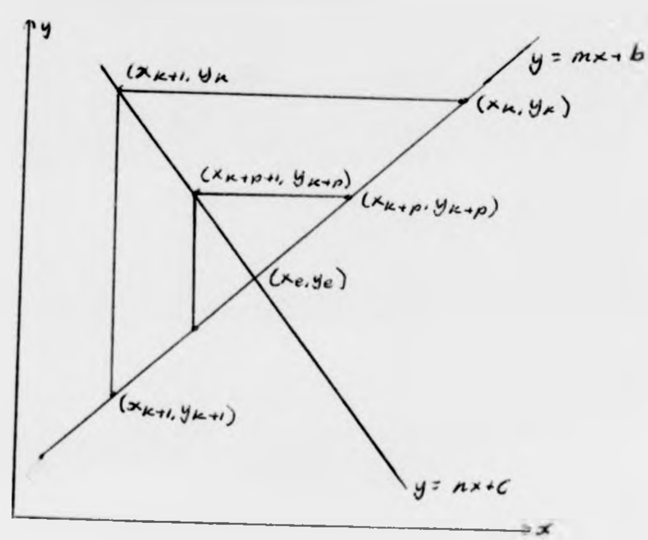


Figure II.19.2

The iteration shown in figure (II.19.2) is specified by

$$y_k = mx_k + b \qquad k = 0,1,2,\ldots$$
$$y_k = nx_{k+1} + c \qquad\qquad (II.19.2)$$

Then $y_k = mx_k + b = nx_{k+1} + c$ or

$$x_{k+1} = \frac{m}{n} x_k + \frac{b-c}{n} \qquad (II.19.3)$$

The general solution of this difference equation is

$$x_k = A\left(\frac{m}{n}\right)^k + \frac{b-c}{n-m} \qquad (II.19.4)$$

where A is an arbitrary constant determined by the value of x_o (see (39) pp 7-27).

The iteration converges if $\lim\limits_{k \to \infty} (x_k - x_e) = 0$ (II.19.5)

Now x_e satisfies $y_e = mx_e + b = nx_e + c$ (II.19.6)

Hence $x_e = \dfrac{b-c}{n-m}$ (II.19.7)

Therefore $x_k - x_e = A\left(\frac{m}{n}\right)^k$ (II.19.8)

and $\lim\limits_{k \to \infty} (x_k - x_e) = \lim\limits_{k \to \infty} A\left(\frac{m}{n}\right)^k$

$$= 0 \text{ if } \left|\frac{m}{n}\right| < 1, \neq 0 \text{ if } \left|\frac{m}{n}\right| \geq 1 \qquad (II.19.9)$$

Thus the anticlockwise iteration converges if $\left|\frac{m}{n}\right| < 1$

(ii) Clockwise Iteration

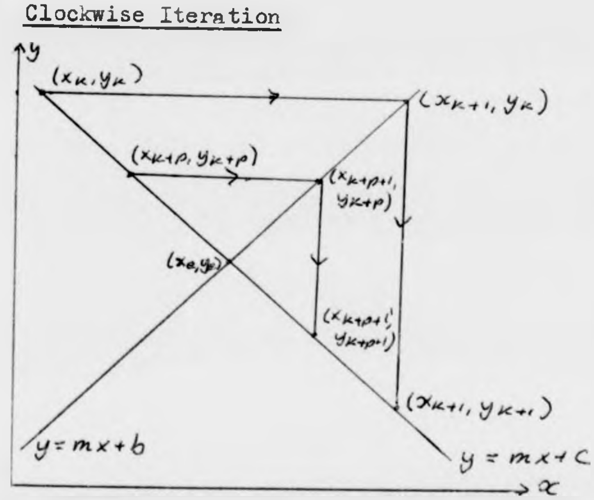


Figure II.19.3

The iteration shown in figure (II.19.3) is specified by

$$y_k = nx_k + c \qquad k = 0,1,2,\ldots \qquad (II.19.10)$$
$$y_k = mx_{k+1} + b$$

Then $y_k = nx_k + c = mx_{k+1} + b$ or

$$x_{k+1} = \frac{n}{m} x_k + \frac{c - b}{m} \qquad (II.19.11)$$

The general solution of equation (II.19.11) is

$$x_k = A \left(\frac{n}{m}\right)^k + \frac{b - c}{n - m} \qquad (II.19.12)$$

Then $\lim_{k \to \infty} (x_k - x_e) = \lim_{k \to \infty} A \left(\frac{n}{m}\right)^k$

$$= 0 \text{ if } \left|\frac{m}{n}\right| > 1, \neq 0 \text{ if } \left|\frac{m}{n}\right| \leq 1 \qquad (II.19.13)$$

Hence the clockwise iteration converges if $\left|\frac{m}{n}\right| > 1$

From equations (II.19.9) and (II.19.13), if $\left|\frac{m}{n}\right| = 1$ there is no convergence.

CHAPTER III

III.1 Introduction

In this chapter an iterative scheme is developed which enables successively better approximations to the upper and lower bounds of the stationary value $L(\phi e, \psi e)$ to be found, where $L(\phi, \psi)$ is a functional. Most of the chapter concentrates on the quadratic functional

$$L(\phi,\psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B\phi\rangle - \tfrac{1}{2}\langle\psi, C\psi\rangle + \langle f,\phi\rangle + \langle g,\psi\rangle \qquad \text{(III.1.1)}$$

which has already been considered several times in Chapter II.

The chapter starts by briefly setting out the dual extremum principles for the functional given in equation (III.1.1); section III.3 considers cobweb iterative schemes for the problem. An explanation of the new iterative scheme is given in section III.4 and section III.5 deals with preliminary considerations.

The next three sections develop the new iterative scheme; section III.6 shows in general how a non-decreasing lower bound sequence can be found, section III.7 gives two lower bound iterative schemes explicitly and section III.8 gives two upper bound iterative schemes. It is also shown in section III.7 by example why it is necessary to have four iterative schemes, compared with the two necessary in cobweb iterations.

Section III.9 considers convergence of the optimizing iterative schemes; two examples are included. The chapter ends with two examples: section III.10 applies the method to the M.H.D. Pipe Flow problem and section III.11 looks at problems specified by the equation $\nabla^2\phi = F'(\phi)$, $F(\phi)$ convex.

CHAPTER III

III.2 Classical Dual Extremum Principles

Let $L(\phi,\psi) : D(L) \subseteq X \times Y \to \mathbb{R}$, where $X \times Y$ is a product space of two vector spaces, be defined by:

$$L(\phi,\psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B\phi\rangle - \tfrac{1}{2}\langle\psi, C\psi\rangle + \langle\phi, f\rangle + \langle\psi, g\rangle \qquad \text{(III.2.1)}$$

$A : D(A) \subseteq Y \to X$ is a linear operator with an adjoint

$A^x : D(A^x) \subseteq X \to Y$ where

$$\langle\phi, A\psi\rangle = \langle\psi, A^x\phi\rangle \quad \forall\,\phi \in D(A) \text{ and } \forall\,\psi \in D(A^x) \qquad \text{(III.2.2)}$$

$B : D(B) \subset X \to X$ and $C : D(C) \subset Y \to Y$ are linear, symmetric, positive-definite operators.

From equations (II.7.11), (II.7.12) the functional derivatives of $L(\phi,\psi)$ are:

$$\nabla_\phi L(\phi,\psi) = A\psi + B\phi + f \qquad \text{(III.2.3)}$$

$$\nabla_\psi L(\phi,\psi) = A^x\phi - C\psi + g \qquad \text{(III.2.4)}$$

From section II.9, $L(\phi,\psi)$ is a strict convex-concave saddle functional; from section II.12 the dual extremum principles are

$$L(\phi_\beta,\psi_\beta) \leq L(\phi e,\psi e) \leq L(\phi_\alpha,\psi_\alpha) \qquad \text{(III.2.5)}$$

where

$$L(\phi_\beta,\psi_\beta) = -\tfrac{1}{2}\langle\phi_\beta, B\phi_\beta\rangle - \tfrac{1}{2}\langle\psi_\beta, C\psi_\beta\rangle + \langle\psi_\beta, g\rangle \qquad \text{(III.2.6)}$$

$$\text{and } A\psi_\beta + B\phi_\beta + f = 0 \qquad \text{(III.2.7)}$$

$$L(\phi_\alpha,\psi_\alpha) = \tfrac{1}{2}\langle\phi_\alpha, B\phi_\alpha\rangle + \tfrac{1}{2}\langle\psi_\alpha, C\psi_\alpha\rangle + \langle\phi_\alpha, f\rangle \qquad \text{(III.2.8)}$$

$$\text{and } A^x\phi_\alpha - C\psi_\alpha + g = 0 \qquad \text{(III.2.9)}$$

$$L(\phi e,\psi e) = \tfrac{1}{2}\langle\phi e, f\rangle + \tfrac{1}{2}\langle\psi e, g\rangle \qquad \text{(III.2.10)}$$

$$\text{and } A\psi e + B\phi e + f = 0,\quad A^x\phi e - C\psi e + g = 0 \qquad \text{(III.2.11)}$$

Finally, from section II.13, as $L(\phi,\psi)$ is a strict saddle functional, if equation (III.2.11) has a solution then it is unique.

CHAPTER III

III.3 Cobweb Iteration for Functionals

As the functional derivatives of $L(\phi,\psi)$, given by equations (III.2.3) and (III.2.4), are linear, cobweb iterative schemes analogous to the one considered for real functions in section II.19 can be set up. At each stage in the iteration a pair of values $(\phi i, \psi j)$ are obtained, and these can be taken as trial functions in the bounds $L(\phi_\beta,\psi_\beta)$ or $L(\phi_\alpha,\psi_\alpha)$, as appropriate. The subscripts of the iterates $(\phi i, \psi j)$ are accommodated in the upper and lower bounds by the use of the following notation:

If $(\phi i, \psi j)$ satisfies $\nabla_\phi L(\phi,\psi) = 0$, then the lower bound found by substituting $(\phi i, \psi j)$ into equation (III.2.6) will be denoted by $\underline{L(\phi i, \psi j)_\beta}$. Similarly, if $(\phi k, \psi l)$ satisfies $\nabla_\psi L(\phi,\psi) = 0$, then the upper bound found by substituting $(\phi k, \psi l)$ into equation (III.2.8) will be denoted by $\underline{L(\phi k, \psi l)_\alpha}$.

Four 'different' cobweb iterative schemes are detailed below. These are 'different' in that the starting point and direction of iteration in each scheme is unique; in fact, as was discussed in section II.19, schemes A and C, and schemes B and D are the same. However, as the method developed in this chapter does produce four different iterative schemes analogous to these four cobweb iterative schemes, all four cobweb iterative schemes will be considered here.

It will be assumed that at each stage in every iteration the given equation has a solution.

Iteration A is defined by the equations:

Choose (ϕ_0,ψ_0) : $\nabla_\phi L(\phi_0,\psi_0) = 0$

Then, for $n = 0,1,2,\ldots$ find $\phi n+1 : \nabla_\psi L(\phi n+1, \psi n) = 0$ (III.3.1)

find $\psi n+1 : \nabla_\phi L(\phi n+1, \psi n+1) = 0$

Figure (III.3.1) illustrates these iterations in a schematic way.

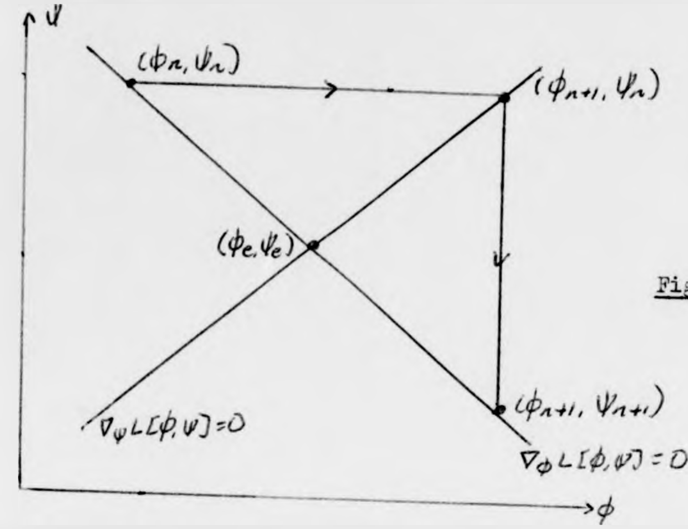


Figure III.3.1

Iteration B is defined by the equations

Choose (ϕ_o, ψ_o) : ∇_ϕ $L(\phi_o, \psi_o) = 0$

Then, for n = 0,1,2,... find $\psi n + 1$: ∇_ψ $L(\phi n, \psi n + 1) = 0$ (III.3.2)

find $\phi n + 1$: ∇_ϕ $L(\phi n + 1, \psi n + 1) = 0$

This iteration is illustrated in figure (III.3.2)

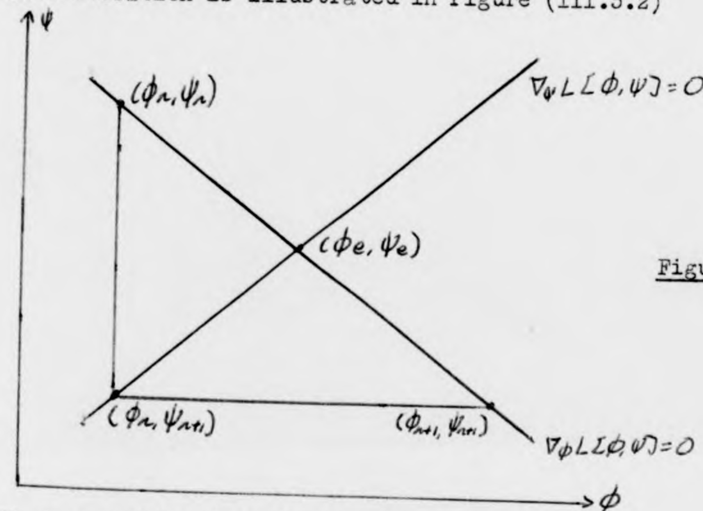


Figure III.3.2

Iteration C is given by the equations

Choose (ϕ_o, ψ_o) : ∇_ψ $L(\phi_o, \psi_o) = 0$

Then, for n = 0,1,2... find $\psi n + 1$: ∇_ϕ $L(\phi n, \psi n + 1) = 0$ (III.3.3)

find $\phi n + 1$: ∇_ψ $L(\phi n + 1, \phi n + 1) = 0$

This iteration is illustrated in figure (III.3.3).

Finally,

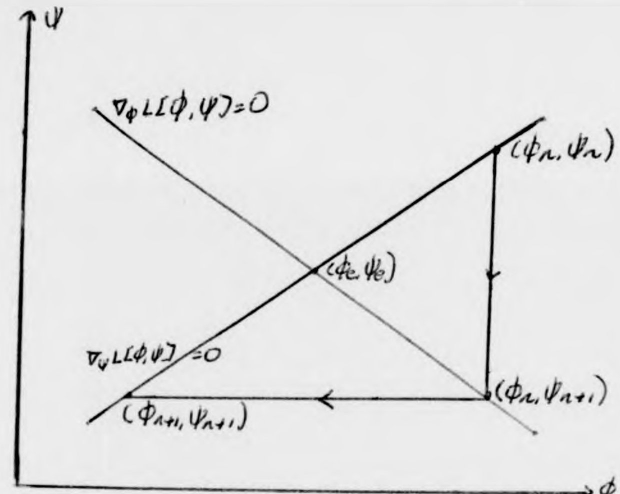
<u>Iteration D</u> is specified by the equations

Choose (ϕ_0, ψ_0) : $\nabla_\psi L(\phi_0, \psi_0) = 0$

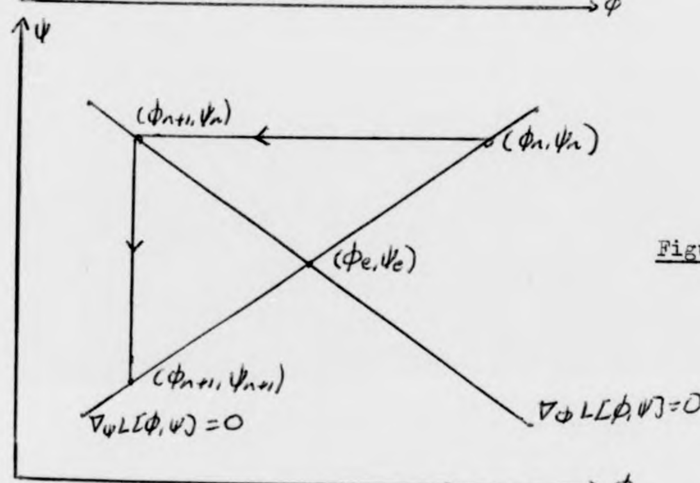
Then, for n = 0,1,2... find ϕ_{n+1} : $\nabla_\phi L(\phi_{n+1}, \psi_n) = 0$ (III.3.4)

find ψ_{n+1} : $\nabla_\psi L(\phi_{n+1}, \psi_{n+1}) = 0$

Figure (III.3.4) illustrates this iteration.



<u>Figure III.3.3</u>



<u>Figure III.3.4</u>

To show that a cobweb iteration converges it is necessary to show either

$$\lim_{n\to\infty} \|\phi_{n+1} - \phi_e\| + \|\psi_{n+1} - \psi_e\| = 0 \qquad \text{(III.3.5)}$$

or, for each iteration in turn,

$$A: \quad \lim_{n\to\infty} \{L(\phi_{n+1}, \psi_n) - L(\phi_n, \psi_n)\} = 0 \qquad \text{(III.3.6)}$$

$$B : \lim_{n\to\infty} \{L(\phi n, \psi n + 1) - L(\phi n, \psi n)\} = 0 \qquad (III.3.7)$$

$$C : \lim_{n\to\infty} \{L(\phi n, \psi n) - L(\phi n, \psi n + 1)\} = 0 \qquad (III.3.8)$$

$$D : \lim_{n\to\infty} \{L(\phi n, \psi n) - L(\phi n + 1, \psi n)\} = 0 \qquad (III.3.9)$$

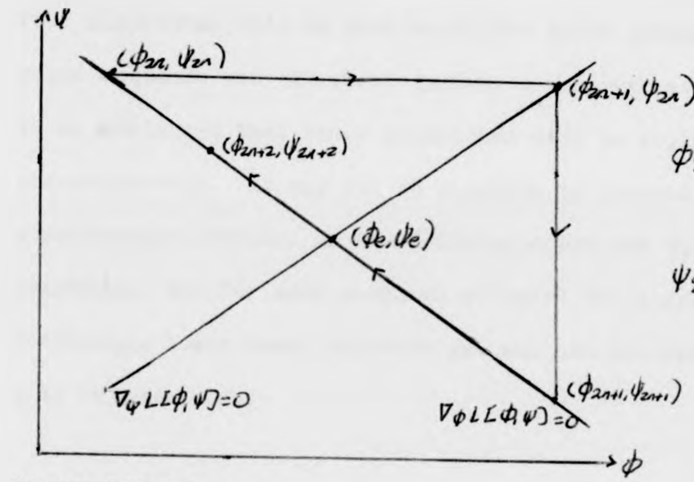
The convergence criteria stated in equation (III.3.5) is stronger than that given in equations (III.3.6) to (III.3.9), in the sense that if equation (III.3.5) is true so is the relevant equation out of (III.3.6) to (III.3.9), depending on which iteration is being considered, but the reverse is not necessarily true (see (74), pp 90 and 112-114).

Now, in section II.19 it was shown, for ordinary functions, that the convergence of a cobweb iterative scheme, when iterating in a particular direction, depends on the relative slopes of the two lines between which the iterations are carried out. Unfortunately, as the equations of the two lines $\nabla_\phi L(\phi,\psi) = 0$ and $\nabla_\psi L(\phi,\psi) = 0$ are operator equations for which the 'slopes' are unknown, we cannot find out using this method, whether the cobweb iterative schemes converge.

In section V.2 conditions are given for which the cobweb iterative schemes do converge; however these conditions on the operators are very restrictive. In the following sections a new iterative scheme is developed which is designed to produce a converging sequence on one or other of the bounds without any restrictions on the operators. Although with this method we only obtain a decreasing upper bound sequence or an increasing lower bound sequence, and we require some restrictions on the operators to prove convergence, we usually find that the optimising iteration does converge.

III.4 Explanation of the Iterative Scheme

Each of the new iterative schemes will follow the same pattern and each one is analogous to one of the cobweb schemes A, B, C and D specified in the previous section. The method is explained by using the new scheme analogous to scheme A; figure (III.4.1) depicts the scheme.



$$\phi_{2n+2} = \lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$$

$$\psi_{2n+2} = \lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$$

Figure III.4.1

(ϕ_0, ψ_0) is chosen to satisfy $\nabla_\phi L(\phi_0, \psi_0) = 0$

Then, for $n = 0,1,2,3,\ldots$ find $\phi_{2n+1} : \nabla_\psi L(\phi_{2n+1}, \psi_{2n}) = 0$

find $\psi_{2n+1} : \nabla_\phi L(\phi_{2n+1}, \psi_{2n+1}) = 0$

The point $(\phi_{2n+2}, \psi_{2n+2})$ is then given by

$$\phi_{2n+2} = \lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$$

$$\psi_{2n+2} = \lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$$

and the iterative cycle is repeated.

$L(\phi_{2n}, \psi_{2n})_\beta$ and $L(\phi_{2n+1}, \psi_{2n})_\alpha$ can be determined at each appropriate point. The method involves using $L(\phi_{2n}, \psi_{2n})_\beta$ to determine a value for λ_{2n+2} which will produce a non-decreasing sequence for the lower bound.

With this method, only one of the bounds in each of the four schemes can be optimized, as there is only one point produced for the other bound during each cycle; in the scheme illustrated in figure (III.4.1) the lower bound can be optimized but not the upper bound. The iteration will, however, produce an upper bound sequence $L(\phi_1,\psi_0)_\alpha$, $L(\phi_3,\psi_2)_\alpha$, ... $L(\phi_{2n+1},\psi_{2n})_\alpha$, which may or may not be a non-increasing sequence. This sequence can be computed during the course of an iteration and the results compared with those obtained by the iterative schemes in which the upper bound is optimized.

Four algorithms will be developed, two which produce a non-decreasing lower bound sequence and two which produce a non-increasing upper bound sequence. It is envisaged that these algorithms will be applied to a problem simultaneously. It may not be possible to proceed with all four algorithms in a particular problem, as the defining equations $\nabla_\phi L = 0$, $\nabla_\psi L = 0$ may not have solutions; but for most problems at least two algorithms, giving one non-decreasing lower bound sequence and one non-increasing upper bound sequence, will be possible.

CHAPTER III

III.5 Preliminary Considerations

In each iterative scheme, ϕ_{2n+2} and ψ_{2n+2} are defined by the equations

$$\phi_{2n+2} = \lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1} \qquad \text{(III.5.1)}$$

and $$\psi_{2n+2} = \lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1} \qquad \text{(III.5.2)}$$

where ϕ_{2n} and ϕ_{2n+1} belong to X, ψ_{2n} and ψ_{2n+1} belong to Y and $\lambda_{2n+2} \in \mathbb{R}$.

A first requirement is that X and Y are affine spaces: that is, using definition (II.8.1),

(i) If ϕ_{2n} and $\phi_{2n+1} \in X$, so does ϕ_{2n+2}; and

(ii) If ψ_{2n} and $\psi_{2n+1} \in Y$, so does ψ_{2n+2}, for all $\lambda_{2n+2} \in \mathbb{R}$.

As $(\phi_{2n+2}, \psi_{2n+2})$ begin a new cycle in each iterative scheme, the second requirement is that wherever the pairs (ϕ_{2n}, ψ_{2n}) and $(\phi_{2n+1}, \psi_{2n+1})$ satisfy $\nabla_\phi L(\phi,\psi) = 0$ or $\nabla_\psi L(\phi,\psi) = 0$, so does the pair $(\phi_{2n+2}, \psi_{2n+2})$.

(i) Let (ϕ_{2n}, ψ_{2n}) and $(\phi_{2n+1}, \psi_{2n+1})$ satisfy $\nabla_\phi L(\phi,\psi) = 0$, then

$$A\psi_{2n} + B\phi_{2n} + f = 0 \qquad \text{(III.5.3)}$$

and $$A\psi_{2n+1} + B\phi_{2n+1} + f = 0 \qquad \text{(III.5.4)}$$

Hence as A and B are linear operators,

$$A(\psi_{2n} - \psi_{2n+1}) + B(\phi_{2n} - \phi_{2n+1}) = 0 \qquad \text{(III.5.5)}$$

Now
$$\begin{aligned}\nabla_\phi L(\phi_{2n+2}, \psi_{2n+2}) &= A\psi_{2n+2} + B\phi_{2n+2} + f \\ &= A(\lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}) \\ &\quad + B(\lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}) + f \\ &= \lambda_{2n+2} (A(\psi_{2n} - \psi_{2n+1}) + B(\phi_{2n} - \phi_{2n+1})) \\ &\quad + A\psi_{2n+1} + B\phi_{2n+1} + f \\ &= 0 \text{ from equations (III.5.4) and (III.5.5).}\end{aligned}$$

Therefore if (ϕ_{2n}, ψ_{2n}) and $(\phi_{2n+1}, \psi_{2n+1})$ satisfy $\nabla_\phi L(\phi,\psi) = 0$ so does $(\phi_{2n+2}, \psi_{2n+2})$.

(ii) Similarly, let (ϕ_{2n}, ψ_{2n}) and $(\phi_{2n+1}, \psi_{2n+1})$ satisfy $\nabla_{\psi} L(\phi,\psi) = 0$: then

$$A^x\phi_{2n} - C\psi_{2n} + g = 0 \qquad \text{(III.5.6)}$$

and $$A^x\phi_{2n+1} - C\psi_{2n+1} + g = 0 \qquad \text{(III.5.7)}$$

As A^x and C are linear operators,

$$A^x(\phi_{2n} - \phi_{2n+1}) - C(\psi_{2n} - \psi_{2n+1}) = 0 \qquad \text{(III.5.8)}$$

$$\begin{aligned}\nabla_{\psi} L(\phi_{2n+2}, \psi_{2n+2}) &= A^x\phi_{2n+2} - C\psi_{2n+2} + g \\ &= A^x(\lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}) \\ &\quad - C(\lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}) + g \\ &= \lambda_{2n+2}(A^x(\phi_{2n} - \phi_{2n+1}) - C(\psi_{2n} - \psi_{2n+1})) \\ &\quad + A^x\phi_{2n+1} - C\psi_{2n+1} + g \\ &= 0 \text{ from equations (III.5.7) and (III.5.8)}\end{aligned}$$

Hence if (ϕ_{2n}, ψ_{2n}) and $(\phi_{2n+1}, \psi_{2n+1})$ satisfy $\nabla_{\psi} L(\phi,\psi) = 0$, so does $(\phi_{2n+2}, \psi_{2n+2})$.

CHAPTER III

III.6 Maximisation of the Lower Bound

In this section the method by which a non-decreasing lower bound sequence can be produced is developed: particular iterations involving the upper bound as well will be considered in the next section.

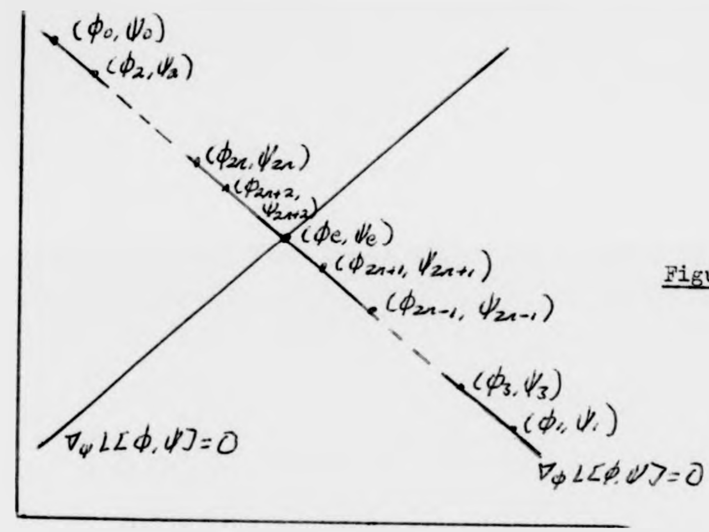


Figure III.6.1

The lower bound iterations will result in a sequence:

$$L(\phi_e, \psi_e) \geq L(\phi_0, \psi_0)_\beta \, , \; L(\phi_1, \psi_1)_\beta \, , \; L(\phi_2, \psi_2)_\beta \, , \; \ldots\ldots$$
$$\ldots\ldots \; L(\phi_{2n}, \psi_{2n})_\beta \, , \; L(\phi_{2n+1}, \psi_{2n+1})_\beta \, , \; L(\phi_{2n+2}, \psi_{2n+2})_\beta \qquad \text{(III.6.1)}$$

where $\phi_{2n+2} = \lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$ (III.6.2)

and $\psi_{2n+2} = \lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$ (III.6.3)

$$\lambda_{2n+2} \in \mathbb{R}$$

λ_{2n+2} needs to be chosen so that a non-decreasing sequence can be obtained from equation (III.6.1). Using equations (III.6.2) and (III.6.3),

$$L(\phi_{2n+2}, \psi_{2n+2})_\beta = L(\lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1},$$
$$\lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1})_\beta \qquad \text{(III.6.4)}$$

$L(\phi_{2n+2}, \psi_{2n+2})_{/\beta}$ is given explicitly for the quadratic functional whose equation is given by equation (III.2.1) in equation (III.2.6):

$$L(\phi_{2n+2}, \psi_{2n+2})_{/\beta} = -\tfrac{1}{2}\langle \phi_{2n+2}, B\phi_{2n+2}\rangle - \tfrac{1}{2}\langle \psi_{2n+2}, C\psi_{2n+2}\rangle + \langle \psi_{2n+2}, g\rangle \qquad \text{(III.6.5)}$$

Hence

$$\begin{aligned} L(\phi_{2n+2}, \psi_{2n+2})_{/\beta} = {} & -\tfrac{1}{2}\langle \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}, \\ & \quad B(\lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1})\rangle \\ & -\tfrac{1}{2}\langle \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}, \\ & \quad C(\lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1})\rangle \\ & + \langle \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}, g\rangle \qquad \text{(III.6.6)} \\ = {} & -\frac{\lambda^2_{2n+2}}{2}\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle \\ & \quad + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\} \\ & - \lambda_{2n+2}\{\langle \phi_{2n} - \phi_{2n+1}, B\phi_{2n+1}\rangle \\ & \quad + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1} - g\rangle\} \\ & - \{\tfrac{1}{2}\langle \phi_{2n+1}, B\phi_{2n+1}\rangle + \tfrac{1}{2}\langle \psi_{2n+1}, C\psi_{2n+1}\rangle \\ & \quad - \langle \psi_{2n+1}, g\rangle\} \qquad \text{(III.6.7)} \end{aligned}$$

using the fact that B and C are symmetric operators.

For the sake of simplicity we let

$$L(\phi_{2n+2}, \psi_{2n+2})_{/\beta} = -\frac{\lambda^2_{2n+2}}{2} P - \lambda_{2n+2} Q - R \qquad \text{(III.6.8)}$$

where P, Q and R are defined by

$$P = \langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle \qquad \text{(III.6.9)}$$

$$Q = \langle \phi_{2n} - \phi_{2n+1}, B\phi_{2n+1}\rangle + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1} - g\rangle \qquad \text{(III.6.10)}$$

and $$R = \tfrac{1}{2}\langle \phi_{2n+1}, B\phi_{2n+1}\rangle + \tfrac{1}{2}\langle \psi_{2n+1}, C\psi_{2n+1}\rangle - \langle \psi_{2n+1}, g\rangle \qquad \text{(III.6.11)}$$

From equation (III.6.8),

$$\frac{dL}{d\lambda_{2n+2}}(\phi_{2n+2}, \psi_{2n+2})_{/\beta} = -\lambda_{2n+2} P - Q \qquad \text{(III.6.12)}$$

$$\frac{d^2L}{d\lambda^2_{2n+2}}(\phi_{2n+2}, \psi_{2n+2})_\beta = -P \qquad \text{(III.6.13)}$$

As B and C are both positive-definite operators, the real number P is always positive and so $\frac{d^2L}{d\lambda^2_{2n+2}}(\phi_{2n+2}, \psi_{2n+2})_\beta$ is always negative; hence $L(\phi_{2n+2}, \psi_{2n+2})_\beta$ is concave in λ_{2n+2}. Therefore the maximum of $L(\phi_{2n+2}, \psi_{2n+2})_\beta$ occurs when $\frac{dL}{d\lambda_{2n+2}}(\phi_{2n+2}, \psi_{2n+2})_\beta$ vanishes.

That is, the maximum occurs when $\lambda_{2n+2} = -\frac{Q}{P}$ (III.6.14)

It should be noted that if P = 0 then Q = 0 giving $\lambda_{2n+2} = 0$ (as P = 0 if and only if $\phi_{2n} - \phi_{2n+1} = 0$ and $\psi_{2n} - \psi_{2n+1} = 0$ because B and C are positive definite operators; then $\phi_{2n} - \phi_{2n+1} = 0$ and $\psi_{2n} - \psi_{2n+1} = 0$ implies that Q = 0), so we do not need to insist that $P \neq 0$.

Substituting equation (III.6.14) into equation (III.6.8) gives

$$\text{Max. } L(\phi_{2n+2}, \psi_{2n+2})_\beta = \frac{Q^2}{2P} - R \qquad \text{(III.6.15)}$$

If we therefore choose $\lambda_{2n+2} = -\frac{Q}{P}$ in equations (III.6.2) and (III.6.3), then

$$L(\phi_{2n+2}, \psi_{2n+2})_\beta \geqslant L(\lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}, \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1})_\beta$$
$$\forall\ \lambda_{2n+2} \in \mathbb{R} \qquad \text{(III.6.16)}$$

In particular, if we take $\lambda_{2n+2} = 0$ on the right hand side of equation (III.6.16),

$$L(\phi_{2n+2}, \psi_{2n+2})_\beta \geqslant L(\phi_{2n+1}, \psi_{2n+1})_\beta \qquad \text{(III.6.17)}$$

and if λ_{2n+2} is taken as 1 on the right hand side of equation (III.6.16),

$$L(\phi_{2n+2}, \psi_{2n+2})_\beta \geqslant L(\phi_{2n}, \psi_{2n})_\beta \qquad \text{(III.6.18)}$$

We therefore have a non-decreasing sequence

$$L(\phi_e, \psi_e) \geqslant L(\phi_{2n+2}, \psi_{2n+2})_\beta \geqslant L(\phi_{2n}, \psi_{2n})_\beta \geqslant \ldots\ldots$$
$$\ldots\ldots \geqslant L(\phi_2, \psi_2)_\beta \geqslant L(\phi_o, \psi_o) \qquad \text{(III.6.19)}$$

Equations (III.6.17) and (III.6.18) will be shown to be true explicitly in the next section, when optimising iterative schemes analogous to schemes A and B in section III.3 are considered.

III.7 Iterative schemes involving maximisation of the Lower Bound

In this section two algorithms are developed for iterative schemes which result in non-decreasing lower bound sequences. These two schemes are analogous to the cobweb iterative schemes A and B in section III.3. By examples, it is shown at the end of the section why both algorithms are necessary.

Iteration A

This is illustrated by Figure (III.7.1) below.

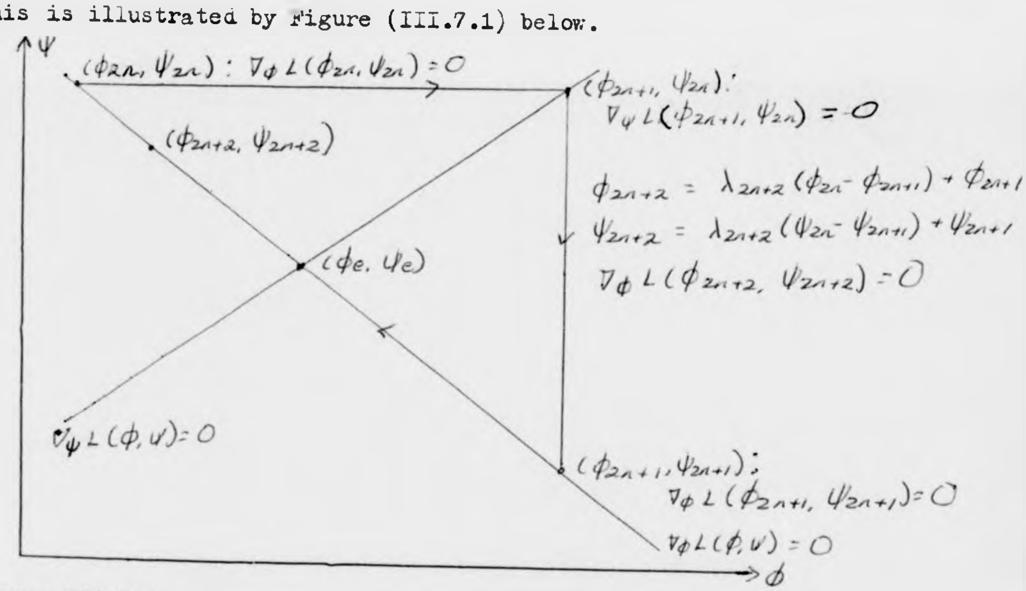


Figure III.7.1

From equations (III.2.3) and (III.2.4), the functional derivatives are given by

$$\nabla_\phi L(\phi_{2n}, \psi_{2n}) = A\psi_{2n} + B\phi_{2n} + f = 0 \qquad \text{(III.7.1)}$$

$$\nabla_\psi L(\phi_{2n+1}, \psi_{2n}) = A^x\phi_{2n+1} - C\psi_{2n} + g = 0 \qquad \text{(III.7.2)}$$

$$\nabla_\phi L(\phi_{2n+1}, \psi_{2n+1}) = A\psi_{2n+1} + B\phi_{2n+1} + f = 0 \qquad \text{(III.7.3)}$$

Before setting out the optimising algorithm, a symmetrical form for Q will be obtained so that the computation required to find λ_{2n+2} is reduced.

From equation (III.6.10),

$$Q = \langle \phi_{2n} - \phi_{2n+1}, B\phi_{2n+1} \rangle + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1} - g \rangle$$

$$= \langle B(\phi_{2n} - \phi_{2n+1}), \phi_{2n+1} \rangle + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1} - g \rangle$$

(as B is a symmetric operator)

$$= -\langle A(\psi_{2n} - \psi_{2n+1}), \phi_{2n+1} \rangle + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1} - g \rangle$$

(using equations (III.7.1) and (III.7.3) and noting that B is a linear operator)

$$= -\langle \psi_{2n} - \psi_{2n+1}, A^x\phi_{2n+1} \rangle + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1} - g \rangle$$

(by equation (III.2.2))

$$= -\langle \psi_{2n} - \psi_{2n+1}, A^x\phi_{2n+1} - C\psi_{2n+1} + g \rangle$$

$$= -\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle \tag{III.7.4}$$

Then $\lambda_{2n+2} = \dfrac{-Q}{P}$ (from equation (III.6.14))

$$= \frac{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle}$$

using equations (III.6.9) and (III.7.4) (III.7.5)

Obviously, as B and C are both positive-definite operators,

$$0 \leq \lambda_{2n+2} \leq 1 \tag{III.7.6}$$

Substituting equations (III.6.9), (III.6.11) and (III.7.4) into equation (III.6.15) gives

$$L(\phi_{2n+2}, \psi_{2n+2})_\beta = \frac{\{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle\}^2}{2\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle\}}$$
$$- \{ \tfrac{1}{2}\langle \phi_{2n+1}, B\phi_{2n+1} \rangle + \tfrac{1}{2}\langle \psi_{2n+1}, C\psi_{2n+1} \rangle - \langle \psi_{2n+1}, g \rangle \} \tag{III.7.7}$$

When computing $L(\phi_{2n+2}, \psi_{2n+2})_\beta$, equation (III.6.5) will probably be more convenient to use than equation (III.7.7); however, equation (III.7.7) can be used to prove equations (III.6.17) and (III.6.18) explicitly.

CHAPTER III

Using equation (III.6.5), equation (III.7.7) can be written as

$$L(\phi_{2n+2}, \psi_{2n+2})_\beta = \frac{\{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}^2}{2\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}} + L(\phi_{2n+1}, \psi_{2n+1})_\beta \qquad \text{(III.7.8)}$$

Hence $L(\phi_{2n+2}, \psi_{2n+2})_\beta - L(\phi_{2n+1}, \psi_{2n+1})_\beta$

$$= \frac{\{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}^2}{2\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}}$$

$\geqslant 0$ as B and C are both positive definite operators.

Therefore $L(\phi_{2n+2}, \psi_{2n+2})_\beta \geqslant L(\phi_{2n+1}, \psi_{2n+1})_\beta$

Using equations (III.6.5) and (III.7.7),

$L(\phi_{2n+2}, \psi_{2n+2})_\beta - L(\phi_{2n}, \psi_{2n})_\beta$

$$= \frac{\{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}^2}{2\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}}$$
$$- \{\tfrac{1}{2}\langle \phi_{2n+1}, B\phi_{2n+1}\rangle + \tfrac{1}{2}\langle \psi_{2n+1}, C\psi_{2n+1}\rangle - \langle \psi_{2n+1}, g\rangle\}$$
$$+ \{\tfrac{1}{2}\langle \phi_{2n}, B\phi_{2n}\rangle + \tfrac{1}{2}\langle \psi_{2n}, C\psi_{2n}\rangle - \langle \psi_{2n}, g\rangle\}$$

$$= \frac{\{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}^2}{2\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}}$$
$$+ \tfrac{1}{2}\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \tfrac{1}{2}\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle$$
$$+ \langle \phi_{2n} - \phi_{2n+1}, B\phi_{2n+1}\rangle + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1} - g\rangle$$

$$= \frac{\{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}^2}{2\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}}$$
$$+ \tfrac{1}{2}\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \tfrac{1}{2}\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle$$
$$- \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle \qquad \text{(from equation (III.7.4))}$$

$$= \frac{\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle\}^2}{2\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}}$$

≥ 0 as B and C are both positive-definite operators

Hence $L(\phi_{2n+2}, \psi_{2n+2})_\beta \geq L(\phi_{2n}, \psi_{2n})_\beta$

The algorithm for iteration A is given as theorem (III.7.1), the algorithm for iteration B will be given immediately after as theorem (III.7.2), with the proof set out in Appendix I.

<u>Theorem (III.7.1)</u>

The iteration specified by the following set of equations:

Choose $(\phi_o, \psi_o) : \nabla_\phi L(\phi_o, \psi_o) = 0 : A\psi_o + B\phi_o + f = 0$

Then, for $n = 0,1,2,3,.....$

Find $\phi_{2n+1} : \nabla_\psi L(\phi_{2n+1}, \psi_{2n}) = 0 : A^x\phi_{2n+1} - C\psi_{2n} + g = 0$

Find $\psi_{2n+1} : \nabla_\phi L(\phi_{2n+1}, \psi_{2n+1}) = 0 : A\psi_{2n+1} + B\phi_{2n+1} + f = 0$

Find $\lambda_{2n+2} = \dfrac{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}$

Let $\phi_{2n+2} = \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$

Let $\psi_{2n+2} = \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$

provides upper and lower bounds to the stationary value $L(\phi e, \psi e)$ which take the forms

$$L(\phi_{2n+1}, \psi_{2n})_\alpha = \tfrac{1}{2}\langle \phi_{2n+1}, B\phi_{2n+1}\rangle + \tfrac{1}{2}\langle \psi_{2n}, C\psi_{2n}\rangle + \langle \phi_{2n+1}, f\rangle$$

and

$$L(\phi_{2n}, \psi_{2n})_\beta = -\tfrac{1}{2}\langle \phi_{2n}, B\phi_{2n}\rangle - \tfrac{1}{2}\langle \psi_{2n}, C\psi_{2n}\rangle + \langle \psi_{2n}, g\rangle$$

The upper and lower bound sequences produced satisfy the following equation:

$$L(\phi_o, \psi_o)_\beta \leq L(\phi_2, \psi_2)_\beta \leq \quad .... \leq L(\phi_{2n}, \psi_{2n})_\beta$$

$$\leq L(\phi_{2n+2}, \psi_{2n+2})_\beta$$

$$\leq L(\phi e, \psi e) \leq$$

$$\text{Min}\,(L(\phi_1, \psi_o)_\alpha,\; L(\phi_3, \psi_2)_\alpha,\; L(\phi_5, \psi_4)_\alpha,\; .....$$

$$..... L(\phi_{2n+1}, \psi_{2n})_\alpha)$$

CHAPTER III

Theorem (III.7.2)

The iteration specified by the following set of equations:

Choose $(\phi_o, \psi_o) : \nabla_\phi L(\phi_o, \psi_o) = 0 : A\psi_o + B\phi_o + f = 0$

Then, for n = 0,1,2,3,...

Find $\psi_{2n+1} : \nabla_\psi L(\phi_{2n}, \psi_{2n+1}) = 0 : A^x\phi_{2n} - C\psi_{2n+1} + g = 0$

Find $\phi_{2n+1} : \nabla_\phi L(\phi_{2n+1}, \psi_{2n+1}) = 0 : A\psi_{2n+1} + B\phi_{2n+1} + f = 0$

Find $\lambda_{2n+2} = \dfrac{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}$

Let $\phi_{2n+2} = \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$

$\psi_{2n+2} = \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$

gives lower and upper bounds to the stationary value $L(\phi_e, \psi_e)$ which are defined by

$$L(\phi_{2n}, \psi_{2n})_\beta = -\tfrac{1}{2}\langle \phi_{2n}, B\phi_{2n}\rangle - \tfrac{1}{2}\langle \psi_{2n}, C\psi_{2n}\rangle + \langle \psi_{2n}, g\rangle$$

and

$$L(\phi_{2n}, \psi_{2n+1})_\alpha = \tfrac{1}{2}\langle \phi_{2n}, B\phi_{2n}\rangle + \tfrac{1}{2}\langle \psi_{2n+1}, C\psi_{2n+1}\rangle + \langle \phi_{2n}, f\rangle$$

The two sequences produced satisfy the following equation:

$$L(\phi_o, \psi_o)_\beta \leq L(\phi_2, \psi_2)_\beta \leq \quad \ldots\ldots \leq L(\phi_{2n}, \psi_{2n})_\beta$$

$$\leq L(\phi_{2n+2}, \psi_{2n+2})_\beta$$

$$\leq L(\phi_e, \psi_e) \leq$$

$$\min\,(L(\phi_o, \psi_1)_\alpha,\ L(\phi_2, \psi_3)_\alpha,\ L(\phi_4, \psi_5)_\alpha,\ \ldots\ldots$$

$$\ldots\ldots L(\phi_{2n}, \psi_{2n+1})_\alpha)$$

Figure (III.7.2) illustrates this iteration.

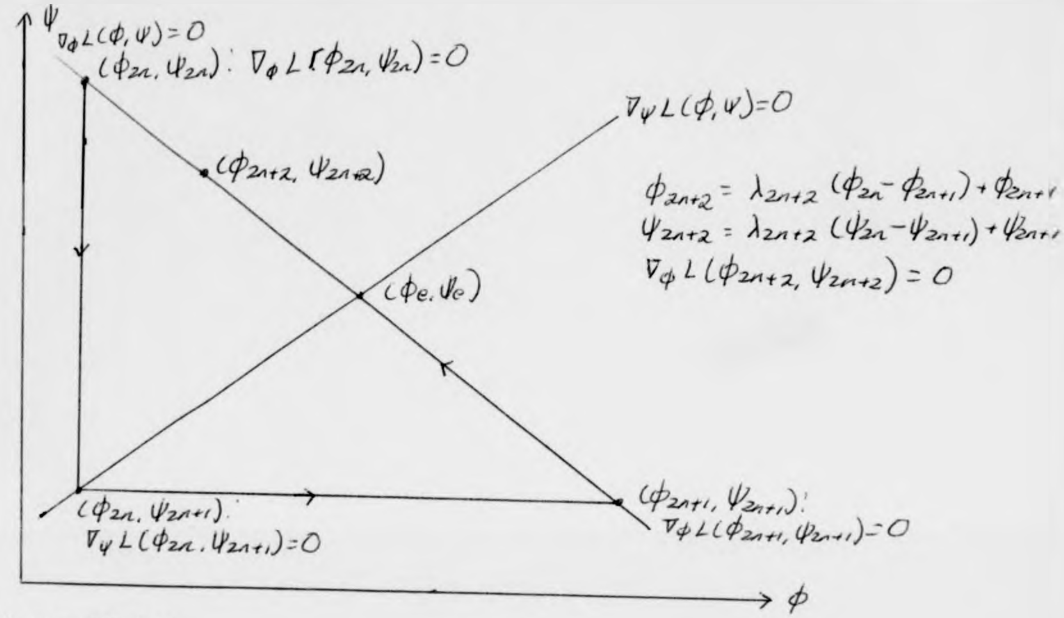


Figure III.7.2

Finally in this section, two examples are provided which justify the inclusion of both theorems (III.7.1) and (III.7.2).

Example 1

From section II.14, the second order ordinary differential equation specified by

$$\frac{d}{dt}\left(r(t)\,\frac{d\phi(t)}{dt}\right) + p(t)\phi(t) + q(t) = 0,\ t\in[a,b],$$

$$p(t) \geq 0 \text{ and } r(t) < 0\ \forall\ t\in[a,b], \qquad \text{(III.7.9)}$$

with boundary conditions $\phi(a) = m$, $\phi'(b) = n$

arises from the quadratic functional $L(\phi,\psi)$ which has the gradients:

$$\nabla_\phi L = \begin{pmatrix} \dfrac{d\psi(t)}{dt} + p(t)\phi(t) + q(t) \\ -\psi(b) + nr(b) \\ 0 \end{pmatrix} \qquad \text{(III.7.10)}$$

and

$$\nabla_{\psi} L = \begin{pmatrix} -\dfrac{d\phi(t)}{dt} + \dfrac{1}{r(t)} \psi(t) \\ 0 \\ -\phi(a) + m \end{pmatrix} \tag{III.7.11}$$

Using these gradients in theorem (III.7.1) gives:

Choose (ϕ_o, ψ_o) : $\dfrac{d\psi_o}{dt}(t) + p(t)\,\phi_o(t) + q(t) = 0,\ \psi_o(b) = nr(b)$ (III.7.12)

Find ϕ_1 : $\dfrac{d\phi_1}{dt}(t) = \dfrac{\psi_o(t)}{r(t)},\quad \phi_1(a) = m$ (III.7.13)

Find ψ_1 : $\dfrac{d\psi_1}{dt}(t) = -p(t)\phi_1(t) - q(t),\ \psi_1(b) = nr(b)$ (III.7.14)

whilst using these gradients in theorem (III.7.2) gives:

Choose (ϕ_o, ψ_o) : $\dfrac{d\psi_o}{dt}(t) + p(t)\,\phi_o(t) + q(t) = 0,\ \psi_o(b) = nr(b)$ (III.7.15)

Find ψ_1 : $\psi_1(t) = r(t)\dfrac{d\phi_o}{dt}(t),\quad \phi_o(a) = m$ (III.7.16)

Find ϕ_1 : $\phi_1(t) = \left[\dfrac{-1}{p(t)}\left(\dfrac{d\psi_1}{dt}(t) + q(t)\right)\right],\ \psi_1(b) = nr(b)$ (III.7.17)

As the iterative steps in equations (III.7.13) and (III.7.14) are both of the form:

Find $u(t)$: $\dfrac{du(t)}{dt} = h(t),\ u(1) = K$, which has a unique solution provided the function $h(t)$ is one which enables the integration to be carried out, the iterations using theorem (III.7.1) can be carried out. However, the iterations using theorem (III.7.2) cannot be carried out; for example, if we let $p(t) = q(t) = 1,\ r(t) = -1,\ a = 0,\ b = 1,\ m = n = 0$, equations (III.7.15) - (III.7.17) become:

Choose (ϕ_o, ψ_o) : $\dfrac{d\psi_o}{dt}(t) + \phi_o(t) + 1 = 0,\ \psi_o(1) = 0$ (III.7.18)

CHAPTER III

Find ψ_1 : $\psi_1(t) = \dfrac{-d\phi_o(t)}{dt}$, $\phi_o(0) = 0$ (III.7.19)

Find ϕ_1 : $\phi_1(t) = -\dfrac{d\psi_1(t)}{dt} + 1$, $\psi_1(1) = 0$ (III.7.20)

Two functions which satisfy equation (III.7.18) and $\phi_o(0) = 0$ are
$\phi_o(t) = -t$, $\psi_o(t) = \dfrac{t^2}{2} - t + \frac{1}{2}$

$\psi_1(t)$ is then required to satisfy equation (III.7.19) and
$\psi_1(1) = 0$ from (III.7.20):
$\psi_1(t) = -1$ using (III.7.19); but $\psi_1(1) = -1$ instead of 0.
The problem in using theorem (III.7.2) arises, of course, because the two conditions: find $\psi_1(t)$ in terms of the derivative of $\phi_o(t)$ and let ψ_1 satisfy a boundary condition, are incompatible.

Hence, for this particular problem, the iterations in theorem (III.7.1) can be used but not those in theorem (III.7.2).

Example 2

The magnetohydrodynamic pipe flow problem described by the pair of equations

$$\left.\begin{aligned} &\nabla^2\phi + M\partial\psi/\partial y = 0 \\ &\nabla^2\psi + M\partial\phi/\partial y + 1 = 0 \end{aligned}\right\} \text{ in } D,\ \phi = \psi = 0 \text{ on } \partial D \qquad (III.7.21)$$

(see equations (I.5.9))

arises from the quadratic function given by the equation

$$L(\phi,\psi) = \int_D \left\{ -M\phi \frac{\partial\psi}{\partial y} - \tfrac{1}{2}\phi\nabla^2\phi + \tfrac{1}{2}\psi\nabla^2\psi + \psi \right\} dD$$

which has the functional derivatives

$$\nabla_\phi L = -M\frac{\partial\psi}{\partial y} - \nabla^2\phi = 0 \text{ in } D,\ \phi = \psi = 0 \text{ on } \partial D \qquad (III.7.22)$$

$$\nabla_\psi L = M\frac{\partial\phi}{\partial y} + \nabla^2\psi + 1 = 0 \text{ in } D,\ \phi = \psi = 0 \text{ on } \partial D \qquad (III.7.23)$$

where ϕ and ψ are both functions of two variables x and y.

Theorem (III.7.1) iterations are then

Choose (ϕ_0, ψ_0) : $M\frac{\partial\psi_0}{\partial y} + \nabla^2\phi_0 = 0$ in D, $\phi_0 = \psi_0 = 0$ on ∂D (III.7.24)

Find ϕ_1 : $M\frac{\partial\phi_1}{\partial y} = -\nabla^2\psi_0 - 1$ in D, $\phi_1 = 0$ on ∂D (III.7.25)

Find ψ_1 : $M\frac{\partial\psi_1}{\partial y} = -\nabla^2\phi_1$ in D, $\psi_1 = 0$ on ∂D (III.7.26)

and theorem (III.7.2) iterations are

Choose (ϕ_0, ψ_0) : $M\frac{\partial\psi_0}{\partial y} + \nabla^2\phi_0 = 0$ in D, $\phi_0 = \psi_0 = 0$ on ∂D (III.7.27)

Find ψ_1: $\nabla^2\psi_1 = -M\frac{\partial\phi_0}{\partial y} - 1$ in D, $\psi_1 = 0$ on ∂D (III.7.28)

Find ϕ_1: $\nabla^2\phi_1 = -M\frac{\partial\psi_1}{\partial y}$ in D, $\phi_1 = 0$ on ∂D (III.7.29)

In principle, equations (III.7.28) and (III.7.29) can be solved, as the function to be found in each case is given in terms of a second order derivative, requiring two integrations for its solution which lead to two variables to be determined.

On the other hand, equations (III.7.25) and (III.7.26) cannot easily be solved, as it is difficult to find functions which satisfy all of the conditions.

Therefore, for this particular problem, the iterations in theorem (III.7.2) can be used, but it is unlikely that those in theorem (III.7.1) could be used.

CHAPTER III

III.8 Iterative schemes involving minimisation of the Upper Bound

This section contains two theorems which set out algorithms for iterations C and D, analogous to cobweb iterations C and D in section III.3. The method used to derive the algorithms is almost the same as that used to obtain the lower bound algorithms in sections III.6 and III.7, the differences arising from the fact that in sections III.6 and III.7 we were finding non-decreasing lower bound sequences, whereas in this section non-increasing upper bound sequences are given. As the methods are so similar, the derivation of the two algorithms is not given here, but is set out briefly in Appendix II.

Theorem III.8.1

The iteration specified by the following set of equations:

Choose (ϕ_0, ψ_0) : $\nabla_\psi L(\phi_0, \psi_0) = 0$: $A^x\phi_0 - C\psi_0 + g = 0$

Then, for n = 0,1,2,3,...

Find ψ_{2n+1}: $\nabla_\phi L(\phi_{2n}, \psi_{2n+1}) = 0$: $A\psi_{2n+1} + B\phi_{2n} + f = 0$

Find ϕ_{2n+1}: $\nabla_\psi L(\phi_{2n+1}, \psi_{2n+1}) = 0$: $A^x\phi_{2n+1} - C\psi_{2n+1} + g = 0$

Find $$\lambda_{2n+2} = \frac{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}$$

Let $\phi_{2n+2} = \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$

Let $\psi_{2n+1} = \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$

provides upper and lower bounds to the stationary value $L(\phi e, \psi e)$ which take the forms

$$L(\phi_{2n}, \psi_{2n})_\alpha = \tfrac{1}{2}\langle \phi_{2n}, B\phi_{2n}\rangle + \tfrac{1}{2}\langle \psi_{2n}, C\psi_{2n}\rangle + \langle \phi_{2n}, f\rangle$$

and

$$L(\phi_{2n}, \psi_{2n+1})_\beta = -\tfrac{1}{2}\langle \phi_{2n}, B\phi_{2n}\rangle - \tfrac{1}{2}\langle \psi_{2n+1}, C\psi_{2n+1}\rangle + \langle \psi_{2n+1}, g\rangle$$

The upper and lower bound sequences produced satisfy the following equation:

$$\text{Max}\ (L(\phi_0, \psi_1)_\beta,\ \ L(\phi_2, \psi_3)_\beta,\ \ldots\ldots\ \ L(\phi_{2n}, \psi_{2n+1})_\beta)$$
$$\leq L(\phi e, \psi e) \leq$$
$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha \leq L(\phi_{2n}, \psi_{2n})_\alpha \leq \ldots\ldots \leq L(\phi_2, \psi_2)_\alpha \leq L(\phi_0, \psi_0)_\alpha$$

This iteration is illustrated in Figure (III.8.1).

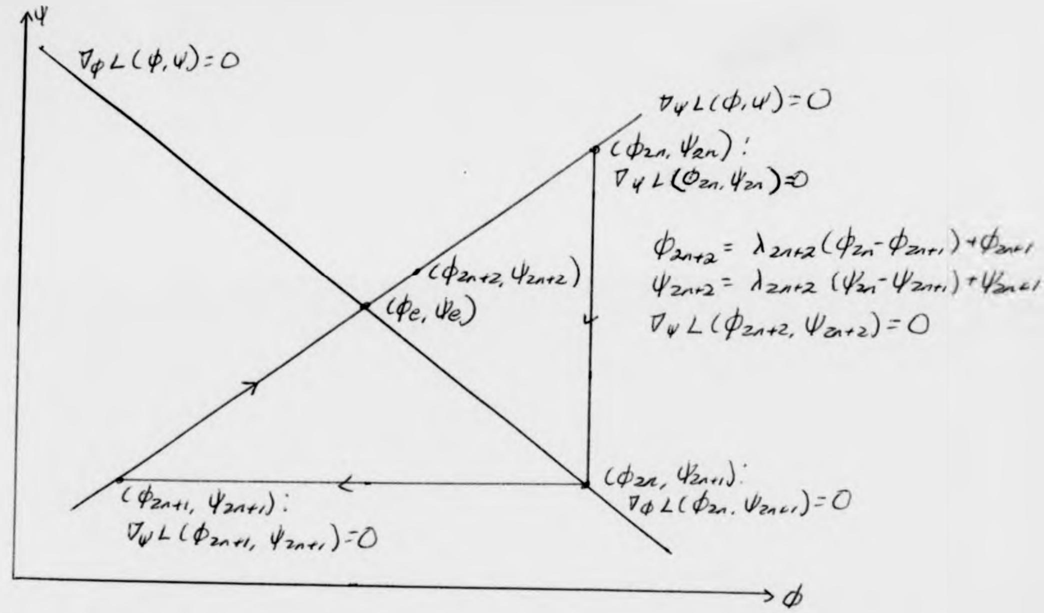


Figure III.8.1

Theorem III.8.2

The iteration specified by the following set of equations:

Choose $(\psi_0, \psi_0) : \nabla_\psi L(\phi_0, \psi_0) = 0 : A^x\phi_0 - C\psi_0 + g = 0$

Then, for $n = 0,1,2,3,\ldots$

Find ϕ_{2n+1}: $\nabla_\phi L(\phi_{2n+1}, \psi_{2n}) = 0 : A\psi_{2n} + B\phi_{2n+1} + f = 0$

Find ψ_{2n+1}: $\nabla_\psi L(\phi_{2n+1}, \psi_{2n+1}) = 0 : A^x\phi_{2n+1} - C\psi_{2n+1} + g = 0$

Find $\lambda_{2n+2} = \dfrac{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}$

Let $\phi_{2n+2} = \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$

$\psi_{2n+2} = \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$

provides upper and lower bounds to the stationary value $L(\phi e, \psi e)$ which take the forms

$$L(\phi_{2n}, \psi_{2n})_\alpha = \tfrac{1}{2}\langle \phi_{2n}, B\phi_{2n}\rangle + \tfrac{1}{2}\langle \psi_{2n}, C\psi_{2n}\rangle + \langle \phi_{2n}, f\rangle$$

and $$L(\phi_{2n+1}, \psi_{2n})_\beta = -\tfrac{1}{2}\langle \phi_{2n+1}, B\phi_{2n+1}\rangle - \tfrac{1}{2}\langle \psi_{2n}, C\psi_{2n}\rangle + \langle \psi_{2n}, g\rangle$$

The upper and lower bound sequences produced satisfy the following equation:

$$\text{Max}\,(L(\phi_1, \psi_0)_\beta,\ L(\phi_3, \psi_2)_\beta, \ldots\ldots\ldots\ldots\ L(\phi_{2n+1}, \psi_{2n})_\beta)$$

$$\leq L(\phi e, \psi e) \leq$$

$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha \leq L(\phi_{2n}, \psi_{2n})_\alpha \leq \ldots \leq L(\phi_2, \psi_2)_\alpha \leq L(\phi_0, \psi_0)_\alpha$$

This iteration is illustrated in Figure (III.8.2).

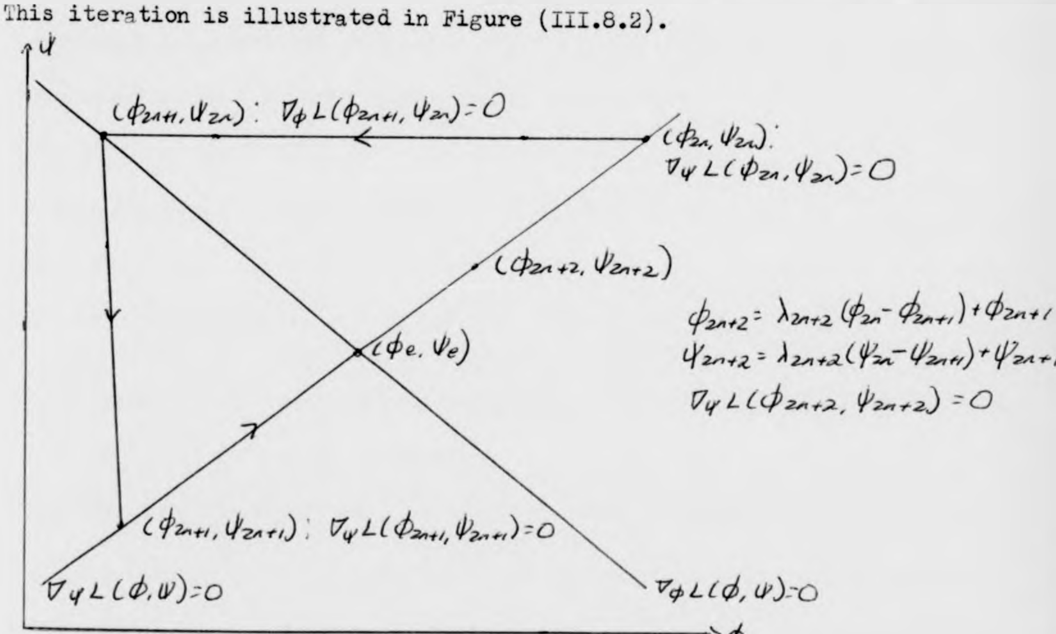


Figure III.8.2

A similar analysis to that in section III.7 shows that there are some problems for which we can use the algorithms given by theorem (III.8.1) but not those given by theorem (III.8.2); and there exist other problems for which the algorithms given by theorem (III.8.2) can be used, but not those given by theorem (III.8.1). For instance, we can use theorem (III.8.1) algorithms but not theorem (III.8.2) algorithms for the second order differential equation problem given as example 1 of the last section; whereas we can use theorem (III.8.2) algorithms for the magnetohydrodynamic pipe flow problem given as example 2 of the last section, but we would have difficulties in using theorem (III.8.2) algorithms for the same problem. Hence, both theorems (III.8.1) and (III.8.2) are necessary.

CHAPTER III

III.9 Convergence of Iterations

It has not been possible to show convergence either to $(\phi e, \psi e)$ or to $L(\phi e, \psi e)$ for the four iterative schemes for general operators A, B and C; in fact, as shown by the examples at the end of section III.7, it may not be possible to use all of the iterations in a particular problem. The flexibility in the schemes, is that the operators A, B and C are unspecified, contributes to the difficulty of proving convergence, since the solution to any operator equation has many solutions.

What is known, however, is that when the iterations can be carried out, converging sequences are produced. The interations given in theorems (III.7.1) and (III.7.2) each produce a non-decreasing sequence:

$$L(\phi_0, \psi_0)_\beta \leq L(\phi_2, \psi_2)_\beta \leq \quad \ldots\ldots\ldots \leq L(\phi_{2n+2}, \psi_{2n+2})_\beta$$

As this sequence is bounded above by $L(\phi e, \psi e)$ the sequence converges to a limit which is a lower bound to $L(\phi e, \psi e)$. This may not be $L(\phi e, \psi e)$; but it will be a better approximation to $L(\phi e, \psi e)$ than is $L(\phi_0, \psi_0)_\beta$.

Similarly, the iterations given in theorems (III.8.1) and (III.8.2) produce a non-increasing monotone sequence:

$$L(\phi_0, \psi_0)_\alpha \geqslant L(\phi_2, \psi_2)_\alpha \geqslant \quad \ldots\ldots\ldots \geqslant L(\phi_{2n+2}, \psi_{2n+2})_\alpha$$

which is bounded below by $L(\phi e, \psi e)$; the sequence therefore converges to a lower bound, which may or may not be $L(\phi e, \psi e)$, but which will be a better approximation to $L(\phi e, \psi e)$ than is $L(\phi_0, \psi_0)_\alpha$

For a problem where at least one of the iterations leading to a monotone upper bound sequence and at least one of the iterations leading to a monotone lower bound sequence can be carried out, the iterations are worked out simultaneously until the difference between $L(\phi_{2n+2}, \psi_{2n+2})_\alpha$ and $L(\phi_{2n+2}, \psi_{2n+2})_\beta$ is less than a given error.

CHAPTER III

For some choices of operators, the iterations specified in the theorems in sections III.7 and III.8 can be shown to converge as the defining equations are those given in one of the methods of steepest descent: this is now considered.

1. Convergence for a bounded operator

By theorem (II.17.1), if P is a linear, self-adjoint, positive-definite operator, then the iteration

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle P\psi_{2n} - G,\ P\psi_{2n} - G\rangle\,(P\psi_{2n} - G)}{\langle P\psi_{2n} - G,\ P(P\psi_{2n} - G)\rangle} \qquad \text{(III.9.1)}$$

converges to the unique solution ψe of $P\psi e = G$ (III.9.2)

Now, from section III.2, the quadratic functional $L(\phi,\psi)$ given in equation (III.2.1) has a stationary value $L(\phi e, \psi e)$ where $(\phi e, \psi e)$ satisfies the equations

$$A\psi e + B\phi e + f = 0 \qquad \text{(III.9.3)}$$

and $A^x\phi e - C\psi e + g = 0$ (III.9.4)

Let us assume that B is bounded below, then by Lemma (II.18.4), B^{-1} exists, and

$$\phi e = - B^{-1}(A\psi e + f) \qquad \text{(III.9.5)}$$

Substituting equation (III.9.5) into (III.9.4) gives

$$- A^x B^{-1}(A\psi e + f) - C\psi e + g = 0$$

or $(A^x B^{-1} A + C)\psi e = g - A^x B^{-1} f$ (III.9.6)

or $P \quad \psi e = G$ (III.9.7)

where $P = A^x B^{-1} A + C$ and $G = g - A^x B^{-1} f$ (III.9.8)

Provided the operator $P = A^x B^{-1} A + C$ is linear, self-adjoint and positive-definite, the iteration given in equation (III.9.1), where $G = g - A^x B^{-1} f$ will converge to the unique solution ψe of equation (III.9.7), assuming that it exists.

CHAPTER III

The purpose of this part of this section is to show that with certain restrictions on the operators A, B and C, the equations in theorems (III.7.1) and (III.7.2) are equivalent to equation (III.9.1); similar results for theorems (III.8.1) and (III.8.2), with ψ replaced by ϕ in equation (III.9.1), will also be given. In all analyses we will assume that $(\phi e, \psi e)$ exists.

Theorem (II.17.1) proves that

$$\lim_{n \to \infty} \| \psi_{2n+2} - \psi_e \| = 0 \qquad \text{(III.9.9)}$$

but what we actually want to prove is that

$$\lim_{n \to \infty} \left\{ \| \phi_{2n+2} - \phi_e \| + \| \psi_{2n+2} - \psi_e \| \right\} = 0 \qquad \text{(III.9.10)}$$

$(\phi_{2n+2}, \psi_{2n+2})$ satisfies the equation

$$A \psi_{2n+2} + B \phi_{2n+2} + f = 0 \qquad \text{(III.9.11)}$$

Using equations (III.9.3) and (III.9.11),

$$\phi_{2n+2} - \phi_e = B^{-1} A (\psi_{2n+2} - \psi_e) \qquad \text{(III.9.12)}$$

If $\lim_{n \to \infty} \| \psi_{2n+2} - \psi_e \| = 0$ then $\lim_{n \to \infty} \| B^{-1} A (\psi_{2n+2} - \psi_e) \| = 0$ and hence $\lim_{n \to \infty} \| \psi_{2n+2} - \psi_e \| = 0$ implies that

$$\lim_{n \to \infty} \| \phi_{2n+2} - \phi_e \| + \| \psi_{2n+2} - \psi_e \| = 0$$

It should also be noted that if equation (III.9.10) is true, then

$$\lim_{n \to \infty} L(\phi_{2n+2}, \psi_{2n+2}) - L(\phi_e, \psi_e) = 0$$

The equations in theorem (III.7.1) can be written

$$\phi_{2n} = - B^{-1} (A \psi_{2n} + f) \qquad \text{(III.9.13)}$$

$$A^x \phi_{2n+1} - C \psi_{2n} + g = 0 \qquad \text{(III.9.14)}$$

$$\phi_{2n+1} = - B^{-1} (A \psi_{2n+1} + f) \qquad \text{(III.9.15)}$$

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle \, (\psi_{2n} - \psi_{2n+1})}{\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle\}} \qquad \text{(III.9.16)}$$

Using equations (III.9.13) and (III.9.15), (III.9.16) can be written

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle \psi_{2n} - \psi_{2n+1}, A^x B^{-1} A(\psi_{2n} - \psi_{2n+1}) \rangle \, (\psi_{2n} - \psi_{2n+1})}{\langle \psi_{2n} - \psi_{2n+1}, (A^x B^{-1} A + C)(\psi_{2n} - \psi_{2n+1}) \rangle} \qquad \text{(III.9.17)}$$

Comparing equations (III.9.1) and (III.9.17), it is obvious that we require $A^x B^{-1} A = I$ and

$P\psi_{2n} - G = \psi_{2n} - \psi_{2n+1}$

or $(A^x B^{-1} A + C)\psi_{2n} - g + A^x B^{-1} f = \psi_{2n} - \psi_{2n+1}$

$(A^x B^{-1} A + C)\psi_{2n} - g + A^x B^{-1} f$

$= A^x B^{-1} (A\psi_{2n} + f) + C\psi_{2n} - g$

$= A^x B^{-1} (A\psi_{2n} + f) + A^x \phi_{2n+1}$ from (III.9.14)

$= A^x B^{-1} (A\psi_{2n} + f) - A^x B^{-1} (A\psi_{2n+1} + f)$ from (III.9.15)

$= A^x B^{-1} A(\psi_{2n} - \psi_{2n+1})$

$= \psi_{2n} - \psi_{2n+1}$ as we require $A^x B^{-1} A = I$.

Letting $A^x B^{-1} A = I$ gives $P = I + C$, which is linear, positive-definite and symmetric. P will be self-adjoint if C is a bounded operator. Therefore, the iterations specified in theorem (III.7.1) will converge to the unique solution ψ_e of equation (III.9.6) provided that the operators A, B and C satisfy the following conditions:

(a) $A^x B^{-1} A = I$

(b) C is a bounded operator.

In a given problem specified by equation (III.9.7), f and g need to be chosen so that $G = g - A^x B^{-1} f$; probably the simplest choice is $g = G$ and $f = 0$.

Similar analyses to the above produce conditions which guarantee that the iterations in theorems (III.7.2), (III.8.1) and (III.8.2) converge; brief details are given in Appendix III and the results are set out below.

The iterations specified in theorem (III.7.2) will converge to the unique solution ψ_e of equation (III.9.6) provided that the operators A, B and C satisfy the following conditions:

(a) $C = I$

(b) $A^x B^{-1} A$ is a linear, self-adjoint operator such that $(A^x B^{-1} A + I)$ is positive-definite.

The iterations specified in theorem (III.8.1) will converge to the unique solution ϕ_e of the equation,

$$(A\, C^{-1} A^x + B)\phi_e = -(A\, C^{-1} g + f) \tag{III.9.18}$$

provided that the operators A, B and C satisfy the following conditions:

(a) $A\, C^{-1} A^x = I$, where C is bounded below

(b) B is a bounded operator.

Finally, the iterations specified in theorem (III.8.2) will converge to the unique solution ϕ_e of equation (III.9.18) provided that the operators A, B and C satisfy the following conditions:

(a) $B = I$

(b) $A\, C^{-1} A^x$ is a linear, self-adjoint operator such that $(A\, C^{-1} A^x + I)$ is positive-definite and C is bounded below.

2. <u>Convergence for an unbounded operator</u>

The purpose of this part of this section is to show that if the operators A, B and C satisfy certain conditions, the iterations in theorems (III.7.1), (III.7.2), (III.8.1) and (III.8.2) converge, because the defining equations are equivalent to those in theorem (II.18.1). For each theorem, conditions on the operators will be found using the equations developed earlier in this section and in Appendix III. The conditions for convergence for theorem (III.7.1) will be developed in this section and the conditions for the other three theorems will be briefly discussed in Appendix III.

By theorem (II.18.1), if P and Q are two closed, symmetric operators with dense domains in a Hilbert space H such that $D(Q) \subseteq D(P)$, and there exist real numbers m, M and Y such that

(a) $\langle x, Qx \rangle \geq Y^2 \langle x, x \rangle \quad \forall x \in D(Q)$ (III.9.19)

and (b) $m \langle x, Qx \rangle \leq \langle x, Px \rangle \leq M \langle x, Qx \rangle \quad \forall x \in D(Q)$ (III.9.20)

then the iteration

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle Z_{2n}, QZ_{2n} \rangle}{\langle Z_{2n}, PZ_{2n} \rangle} Z_{2n} \qquad \text{(III.9.21)}$$

with Z_{2n} given by $QZ_{2n} = P\psi_{2n} - G$ (III.9.22)

converges to the unique solution ψ_e of $P\psi_e = G$ (III.9.23)

We are going to assume that (ϕ_e, ψ_e) exists.

From equation (III.9.8), $P = A^x B^{-1} A + C$ and $G = g - A^x B^{-1} f$ (III.9.24)

(where we assume that B is bounded below).

From equation (III.9.17), the iteration in theorem (III.7.1) is

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle \psi_{2n} - \psi_{2n+1}, A^x B^{-1} A(\psi_{2n} - \psi_{2n+1}) \rangle (\psi_{2n} - \psi_{2n+1})}{\langle \psi_{2n} - \psi_{2n+1}, (A^x B^{-1} A + C)(\psi_{2n} - \psi_{2n+1}) \rangle} \qquad \text{(III.9.25)}$$

Comparing equations (III.9.21) and (III.9.25), it can be seen that we must define Z_{2n} and Q as

$$Z_{2n} = \psi_{2n} - \psi_{2n+1} \qquad \text{(III.9.26)}$$

$$Q = A^x B^{-1} A \qquad \text{(III.9.27)}$$

Then, to satisfy equation (III.9.22), the iteration equations (III.9.13) - (III.9.15) must be equivalent to

$$A^x B^{-1} A(\psi_{2n} - \psi_{2n+1}) = (A^x B^{-1} A + C)\psi_{2n} + A^x B^{-1} f - g$$

or $A^x B^{-1} (A\psi_{2n+1} + f) + C\psi_{2n} - g = 0$ (III.9.28)

From equations (III.9.14) and (III.9.15),

$A^x B^{-1} (A \psi_{2n+1} + f) = - A^x \phi_{2n+1} = -(C \psi_{2n} - g)$ and hence the iteration equations (III.9.13) - (III.9.15) are equivalent to equation (III.9.22). Therefore, the iteration given in theorem (III.7.1) will converge to the solution ψ_e of

$$(A^x B^{-1} A + C)\psi_e = g - A^x B^{-1} f \qquad \text{(III.9.29)}$$

provided

(a) $A^x B^{-1} A$ and C are closed symmetric operators with dense domains on a Hilbert space H, with

$D(A^x B^{-1} A) \subseteq D(A^x B^{-1} A + C)$

(b) There exists a real number Y such that

$$\langle x, A^x B^{-1} Ax \rangle \geq Y^2 \langle x,x \rangle \quad \forall\, x \in D(A^x B^{-1} A) \qquad \text{(III.9.30)}$$

(c) There exist real numbers m and M such that

$$(m-1)\langle x, A^x B^{-1} Ax \rangle \leq \langle x, Cx \rangle \leq (M-1)\langle x, A^x B^{-1} Ax \rangle \quad \forall\, x \in D(A^x B^{-1} A) \qquad \text{(III.9.31)}$$

The iteration given in theorem (II.7.2) will converge to the solution ψ_e of equation (III.9.29), provided

(a) $A^x B^{-1} A$ and C are closed, symmetric operators with dense domains on a Hilbert space H, with $D(C) \subseteq D(A^x B^{-1} A + C)$

(b) There exists a real number Y such that

$$\langle x, Cx \rangle \geq Y^2 \langle x,x \rangle \quad \forall\, x \in D(C) \qquad \text{(III.9.32)}$$

(c) There exists real numbers m and M such that

$$(m-1)\langle x, Cx \rangle \leq \langle x, A^x B^{-1} Ax \rangle \leq (M-1)\langle x, Cx \rangle \quad \forall\, x \in D(C) \qquad \text{(III.9.33)}$$

The iteration given in theorem (III.8.1) will converge to the unique solution ϕ_e of

$$(AC^{-1} A^x + B)\phi_e = -(AC^{-1} g + f) \qquad \text{(III.9.34)}$$

(where we assume that C is bounded below), provided that

(a) $AC^{-1}A^x$ and B are closed, symmetric operators with dense domains on a Hilbert space H such that $D(AC^{-1}A^x) \subseteq D(AC^{-1}A^x + B)$

(b) There exists a real number Y such that

$$\langle x, AC^{-1}A^x x\rangle \leq Y^2\langle x,x\rangle \quad \forall x \in D(AC^{-1}A^x) \qquad \text{(III.9.35)}$$

(c) There exist real numbers m and M such that

$$(m-1)\langle x, AC^{-1}A^x x\rangle \leq \langle x, Bx\rangle \leq (M-1)\langle x, AC^{-1}A^x x\rangle \quad \forall x \in D(AC^{-1}Ax) \qquad \text{(III.9.36)}$$

Finally, the iteration given in theorem (III.8.2) will converge to the unique solution ϕ_e of equation (III.9.34), provided

(a) $AC^{-1}A^x$ and B are closed, symmetric operators with $D(B) \subseteq D(AC^{-1}A^x + B)$

(b) There exists a real number Y such that

$$\langle x, Bx\rangle \geq Y^2\langle x,x\rangle \quad \forall x \in D(B) \qquad \text{(III.9.37)}$$

(c) There exist real numbers m and M such that

$$(m-1)\langle x, Bx\rangle \leq \langle x, AC^{-1}A^x x\rangle \leq (M-1)\langle x, Bx\rangle \quad \forall x \in D(B) \qquad \text{(III.9.38)}$$

Two simple examples are given as illustrations. Example 1 considers operators which satisfy the conditions for convergence for bounded operators and example 2 considers operators which satisfy the conditions for convergence for unbounded operators.

Example 1

We consider the problem

$$(K + I)\phi_e(x) = x^2 - 2x, \quad x \in [0,1] \qquad \text{(III.9.39)}$$

where $K\phi(x) = \int_0^1 k(x,y)\,\phi(y)\,dy$ (III.9.40)

and $k(x,y) = \begin{cases} y, & x \geq y \\ x, & x \leq y \end{cases} \quad x,y \in [0,1]$ (III.9.41)

CHAPTER III

From example 1 of section II.15, the operator K is linear, self-adjoint and positive-definite, and is therefore a bounded operator. The conditions for convergence for the iteration given by theorem (III.8.1) for bounded operators (see page 156) will be satisfied if we take

$$A = A^x = C = I, \quad B = K, \quad f = 2x - x^2, \quad g = 0 \tag{III.9.42}$$

Then theorem (III.8.1) iterations reduce to:

Choose ϕ_o arbitrarily:

For $n = 0,1,2,\ldots$ Find ϕ_{2n+1}: $\phi_{2n+1} = x^2 - 2x - K\phi_{2n}$

$$\text{Find } \lambda_{2n+2} = \frac{\langle \phi_{2n} - \phi_{2n+1}, K(\phi_{2n} - \phi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, (K+I)(\phi_{2n} - \phi_{2n+1})\rangle} \tag{III.9.43}$$

Let $\phi_{2n+2} = \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$

$L(\phi_{2n}, \psi_{2n})_\alpha = \frac{1}{2}\langle \phi_{2n}, (K+I)\phi_{2n} + 2(2x - x^2)\rangle$

$L(\phi_{2n}, \psi_{2n+1})_\beta = -\frac{1}{2}\langle \phi_{2n}, K\phi_{2n}\rangle - \frac{1}{2}\langle \phi_{2n+1}, \phi_{2n+1}\rangle$

The inner product is given by the equation

$$\langle h_1(x), h_2(x)\rangle = \int_0^1 h_1(x)\, h_2(x)\, dx \tag{III.9.44}$$

The first few iterations starting with $\phi_o = 0$ are set out in Table (III.9.1); the working is given in Appendix IV.

n	ϕ_{2n}	ϕ_{2n+1}	λ_{2n+2}	$L(\phi_{2n}, \psi_{2n+1})_\beta$	$L(\phi_{2n}, \psi_{2n})_\alpha$
0	0	$-2x + x^2$	$\frac{17}{59}$	$-0.26\dot{6}$	0
1	$\frac{-84}{59}x + \frac{42}{59}x^2$	$\frac{1}{118}(-180x + 118x^2 - 28x^3 + 7x^4)$	$\frac{101}{2411}$	-0.18985604	-0.1898305
2	$-1.5212x + 0.9879x^2 - 0.2273x^3 + 0.05684x^4$				-0.18985497

Table (III.9.1)

By converting equation (III.9.39) to its equivalent differential equation and solving, we find that $\phi_e(x) = 2\cosh x - 2\tanh 1 \sinh x - 2$, and

$L(\phi_e, \psi_e) = \frac{4}{3} - 2 \tanh 1 = -0.18985497$. The first four terms in $\phi_e(x)$ are $-1.5232\, x + x^2 - 0.2539\, x^3 + 0.05684\, x^4$; as expected, the iterations seem to be converging both to $L(\phi_e, \psi_e)$ and ϕ_e.

Example 2

Consider the problem $\phi''(t) - \phi(t) = 2,\ t \in [0,1]$,

$$\phi(0) = \phi'(1) = 0 \qquad \text{(III.9.45)}$$

From section II.14, this problem arises from the derivative of the convex/concave saddle functional

$$L(\phi, \psi) = \int_0^1 \left\{ \phi(t) \frac{d\psi(t)}{dt} + \tfrac{1}{2} (\phi(t))^2 - \tfrac{1}{2} (\psi(t))^2 + 2\phi(t) \right\} dt$$

$$- \phi(1)\, \psi(1) \qquad \text{(III.9.46)}$$

To use the convergence analysis given earlier in this section on page 158, we need to specify operators A, A^x, B and C. Using the inner product from section II.14,

$$\langle r,s \rangle = \int_0^1 r(t)\, s(t)\, dt + r(b)\, s(b) + r(a)\, s(a),$$

we can specify these operators as

$$A = \begin{pmatrix} \frac{d}{dt} \text{ in }]0,1[\\ -I \text{ at } t = 1 \\ 0 \text{ at } t = 0 \end{pmatrix} \qquad A^x = \begin{pmatrix} -\frac{d}{dt} \text{ in }]0,1[\\ 0 \text{ at } t = 1 \\ -I \text{ at } t = 0 \end{pmatrix}$$

$$B = C = \begin{pmatrix} I \text{ in }]0,1[\\ 0 \text{ at } t = 1 \\ 0 \text{ at } t = d \end{pmatrix} \qquad \text{(III.9.47)}$$

Then we have
$$AC^{-1} A^x = \begin{pmatrix} -\frac{d^2}{dt^2} \text{ in }]0,1[\\ 0 \text{ at } t = 1 \\ 0 \text{ at } t = 0 \end{pmatrix} \qquad \text{(III.9.48)}$$

$$\text{and } AC^{-1}A^x + B = \begin{pmatrix} -\dfrac{d^2}{dt^2} + I \text{ in }]0,1[\\ 0 \text{ at } t = 1 \\ 0 \text{ at } t = 0 \end{pmatrix} \qquad \text{(III.9.49)}$$

From page 158, the iteration specified in theorem (III.8.1) will converge to the unique solution of equation (III.9.45) provided

(a) $AC^{-1}A^x$ and B are closed symmetric operators with dense domains on a Hilbert space H such that $D(AC^{-1}A^x) \subseteq D(AC^{-1}A^x + B)$.

We can take the domains of $AC^{-1}A^x$ and B as $C^2(0,1)$, the space of twice continuously differentiable functions on $]0,1[$; then by definition (II.5.11) and the example which follows it, $AC^{-1}A^x$ and B are closed operators. It can be shown using definition (II.5.12) that $C^2(0,1)$ is a dense domain, and by using equation (II.5.6) that $AC^{-1}A^x$ and $AC^{-1}A^x + B$ are both symmetric operators; hence this condition is satisfied.

(b) There exists a real number Y such that

$$\langle\phi, AC^{-1}A^x\phi\rangle \geq Y^2\langle\phi,\phi\rangle \quad \forall\phi \in D(AC^{-1}A^x)$$

That is, we require there to exist $Y \in \mathbb{R}$ such that

$$\int_0^1 -\frac{d^2\phi}{dt^2}(t)\,dt \geqslant Y^2\int_0^1 (\phi(t))^2\,dt \qquad \text{(III.9.50)}$$

As $\phi(0) = \phi'(1) = 0$, the left hand side of equation (III.9.50) is equivalent to $\int_0^1 \left(\frac{d\phi}{dt}(t)\right)^2 dt$; hence we require $Y \in \mathbb{R}$ such that

$$\int_0^1 (\phi(t)^2\,dt \leqslant \frac{1}{Y^2}\int_0^1 \left(\frac{d\phi}{dt}(t)\right)^2 dt \qquad \text{(III.9.51)}$$

Now, by Lemma (II.4.2),

$$\int_0^1 (\phi(t)^2\,dt \leqslant \tfrac{1}{2}\int_0^1 \left(\frac{d\phi}{dt}(t)\right)^2$$

Therefore, one such Y which satisfies the condition is $Y = \sqrt{2}$, and hence the condition is satisfied.

(c) There exist real numbers m and M such that

$$(m-1)\langle\phi,\ AC^{-1}\ A^x\phi\rangle \leq \langle\phi,\ B\phi\rangle \leq \ (M-1)\langle\phi,\ AC^{-1}\ A^x\phi\rangle$$
$$\forall\phi\in\ D(AC^{-1}\ A^x)$$

That is, we require that there exist m and $M \in \mathbb{R}$ such that

$$(m-1)\int_0^1\left(\frac{d\phi(t)}{dt}\right)^2 dt \leq \int_0^1 (\phi(t))^2\ dt \leq\ (M-1)\int_0^1\left(\frac{d\phi(t)}{dt}\right)^2 dt \qquad \text{(III.9.52)}$$

As the integrals $\int_0^1\left(\frac{d\phi(t)}{dt}\right)^2 dt$ and $\int_0^1 (\phi(t))^2\ dt$ are greater than or equal to zero, the left hand inequality is satisfied as we can take $m \leq 1$, making

$$(m-1)\int_0^1\left(\frac{d\phi(t)}{dt}\right)^2 dt \leq\ 0 \leq \int_0^1 (\phi(t))^2\ dt$$

The right hand inequality is satisfied if we take $M \geqslant \frac{3}{2}$, for then $M - 1 \geqslant \frac{1}{2}$ and hence, using equation (III.9.51),

$$\int_0^1 (\phi(t))^2\ dt \ \leq\ \tfrac{1}{2}\int_0^1\left(\frac{d\phi(t)}{dt}\right)^2 dt \leq (M-1)\int_0^1\left(\frac{d\phi(t)}{dt}\right)^2 dt$$

Therefore, this condition is satisfied, and so all of the conditions for convergence for the iteration given by theorem (III.8.1), for unbounded operators, are satisfied.

For this problem, theorem (III.8.1) iterations are:

Choose (ϕ_0, ψ_0) : $\dfrac{d\phi_0}{dt} + \psi_0\ = C,\ \phi_0\ (C) = 0$

Then, for n = 0,1,2,3,... find ψ_{2n+1}: $\dfrac{d\psi_{2n+1}}{dt} + \phi_{2n+2} = 0$
$\psi_{2n+1}\ (1) = 0$

Find ϕ_{2n+1}: $\dfrac{d\phi_{2n+1}}{dt} + \psi_{2n+1} = 0,\ \phi_{2n+1}\ (0) = 0 \qquad \text{(III.9.52)}$

CHAPTER III

Find $$\lambda_{2n+2} = \frac{\int_0^1 (\phi_{2n} - \phi_{2n+1})^2 \, dt}{\int_0^1 \{(\phi_{2n} - \phi_{2n+1})^2 + (\psi_{2n} - \psi_{2n+1})^2\} \, dt}$$

Let $$\phi_{2n+2} = \lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$$

$$\psi_{2n+2} = \lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1} \qquad \text{(III.9.52)}$$

$$L(\phi_{2n}, \psi_{2n})_\alpha = \tfrac{1}{2}\int_0^1 \{\phi_{2n}^2 + \psi_{2n}^2 + 4\phi_{2n}\} \, dt$$

$$L(\phi_{2n}, \psi_{2n+1})_\beta = -\tfrac{1}{2}\int_0^1 \{\phi_{2n}^2 + \psi_{2n+1}^2\} \, dt$$

The first few iterations starting with $\phi_0 = 0$ are set out in Table (III.9.1), with details given in Appendix IV. As the 'ψ' terms are only intermediate terms, these have been omitted from the table for clarity.

n	ϕ_{2n}	ϕ_{2n+1}	λ_{2n+2}	$L(\phi_{2n}, \psi_{2n+1})_\beta$	$L(\phi_{2n}, \psi_{2n})_\alpha$
0	0	$t^2 - 2t$	$\frac{2}{7}$		0
1	$\frac{5}{7}(-2t + t^2)$	$-\frac{32}{21}t + t^2 - \frac{5}{21}t^3 + \frac{5}{84}t^4$	$\frac{445883}{449063}$	-0.476838354	-0.476190476
2	$-1.4786t + 0.7144t^2 - 0.00009544t^3 + 0.00002386t^4$				-0.47619099

Table (III.9.2)

From Example 1, the first four terms in $\phi_e(t)$ are $-1.5232t + t^2 - 0.2539t^3 + 0.0833t^4$ and from Appendix IV, $L(\phi_e, \psi_e) = -0.47681168$.

As in example 1, the iterations seem to be converging to both $L(\phi_e, \psi_e)$ and ϕ_e; the convergence is faster in example 1 than in example 2, but a lot more work is necessary to carry out the iterations in example 1 than in example 2.

CHAPTER III

III.10 Application of the Optimising Iteration to the Magnetohydrodynamic Pipe Flow Problem

The M.H.D. problem, which has been briefly considered earlier in this chapter, in section III.7, is specified by the pair of equations

$$\left.\begin{aligned} &M\frac{\partial\psi}{\partial y} + \nabla^2\phi = 0 \text{ in } D \\ &M\frac{\partial\phi}{\partial y} + \nabla^2\psi + 1 = 0 \text{ in } D \end{aligned}\right\} \quad \phi = \psi = 0 \text{ on } \partial D \qquad \text{(III.10.1)}$$

In this section it is shown that the optimising iteration developed in this chapter, when applied in a particular way to the M.H.D. problem, converges without any restrictions on M, unlike the cobweb iteration given in (66).

Comparing equation (III.10.1) with equations (III.2.3) and (III.2.4), it is seen that we can take

$$A = -\frac{M\partial}{\partial y}, \quad A^x = \frac{M\partial}{\partial y}, \quad B = -\nabla^2, \quad C = -\nabla^2, \quad f = 0 \text{ and } g = 1 \qquad \text{(III.10.2)}$$

with the assumption that $\phi = \psi = 0$ on the boundary.

With the usual inner product,

$$L(\phi, \psi) = \int_D \left\{ -\phi M\frac{\partial\psi}{\partial y} - \tfrac{1}{2}\phi\nabla^2\phi + \tfrac{1}{2}\psi\nabla^2\psi + \psi \right\} dD \qquad \text{(III.10.3)}$$

which is, as required, a convex/concave saddle functional as $-\nabla^2$ is a positive-definite operator. The gradients of $L(\phi, \psi)$ are given by

$$\left.\begin{aligned} &\nabla_\phi L(\phi, \psi) = -M\frac{\partial\psi}{\partial y} - \nabla^2\phi \\ &\nabla_\psi L(\phi, \psi) = M\frac{\partial\phi}{\partial y} + \nabla^2\psi + 1 \end{aligned}\right. \qquad \text{(III.10.4)}$$

Setting equation (III.10.4) equal to zero gives the M.H.D. problem, (with the assumption that $\phi = \psi = 0$ on the boundary).

For this particular functional, cobweb iteration B, equation (III.3.2), is written:

Choose (ψ_0) : $\psi_0 = 0$ on the boundary

For n = 0,1,2,... : find ϕ_n : $\nabla^2\phi_n = -M\frac{\partial\psi_n}{\partial y}$, $\phi_n = 0$ on the boundary

find ψ_{n+1}: $\nabla^2\psi_{n+1} = -M\frac{\partial\phi_n}{\partial y} - 1$, $\psi_{n+1} = 0$ on the boundary

(III.10.5)

Using equations (III.10.2), (III.2.6) and (III.2.8), the lower and upper bounds are:

$$L(\phi_n, \psi_n)_\beta = \int_D \left\{ \tfrac{1}{2}\phi_n \nabla^2\phi_n + \tfrac{1}{2}\psi_n\nabla^2\psi_n + \psi_n \right\} dD \qquad \text{(III.10.6)}$$

$$L(\phi_n, \psi_{n+1})_\alpha = \int_D \left\{ -\tfrac{1}{2}\phi_n \nabla^2\phi_n - \tfrac{1}{2}\psi_{n+1}\nabla^2\psi_{n+1} \right\} dD \qquad \text{(III.10.7)}$$

Using equations (III.10.5), (III.10.6) and (III.10.7) can be written:

$$L(\phi_n, \psi_n)_\beta = \tfrac{1}{2}\int_D \left\{ \psi_n - M\psi_n \frac{\partial}{\partial y}(\phi_{n-1} - \phi_n) \right\} dD \qquad \text{(III.10.8)}$$

$$L(\phi_n, \psi_{n+1})_\alpha = \tfrac{1}{2}\int_D \left\{ \psi_{n+1} - M\phi_n \frac{\partial}{\partial y}(\psi_{n+1} - \psi_n) \right\} dD \qquad \text{(III.10.9)}$$

and in (66) it is shown that

$\lim_{n\to\infty} L(\phi_n, \psi_{n+1})_\alpha - L(\phi_n, \psi_n)_\beta = 0$ provided M satisfies the equation

$$M < \sqrt{\mu} \text{ where } \frac{1}{\mu} < a^2 \qquad \text{(III.10.10)}$$

and a is the width of D, projected onto the x-axis.

The equivalent optimising iteration for the M.H.D. problem given by equations (III.10.1) is that given in theorem (III.7.2) ∇^2 is an unbounded operator, and therefore there are two possible methods of showing convergence: one is similar to the method used above for the cobweb iteration and the other is the method given in the previous section. This latter method cannot be used as the inverse of the operator $B = -\nabla^2$ cannot be easily found.

Theorem (III.7.2) results in the bounds

$$L(\phi_{2n}, \psi_{2n})_{\beta} = \int_D \left\{ \tfrac{1}{2}\phi_{2n} \nabla^2 \phi_{2n} + \tfrac{1}{2} \psi_{2n} \nabla^2 \psi_{2n} + \psi_{2n} \right\} dD$$

and

$$L(\phi_{2n}, \psi_{2n+1})_{\alpha} = \int_D \left\{ -\tfrac{1}{2} \phi_{2n} \nabla^2 \phi_{2n} - \tfrac{1}{2} \psi_{2n+1} \nabla^2 \psi_{2n+1} \right\} dD$$

where $M\dfrac{\partial \psi_{2n}}{\partial y} + \nabla^2 \phi_{2n} = 0$

$$M\frac{\partial \phi_{2n}}{\partial y} + \nabla^2 \psi_{2n+1} + 1 = 0$$

$$M\frac{\partial \psi_{2n+1}}{\partial y} + \nabla^2 \phi_{2n+1} = 0$$

$$\phi_{2n+2} = \lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$$

$$\psi_{2n+2} = \lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$$

The difference $L(\phi_{2n}, \psi_{2n+1})_{\alpha} - L(\phi_{2n}, \psi_{2n})_{\beta}$ cannot be simplified in the same way as the same difference in the cobweb iteration, as the simplification would involve the equation $M\dfrac{\partial \phi_{2n-1}}{\partial y} + \nabla^2 \phi_{2n+1} = 0$

which does not appear in the equations for the optimising iteration; I have not been able to prove that the optimising iteration converges by any other method, for the M.H.D. problem given by equations (III.10.1).

However, by uncoupling the equations, convergence can be proven, and this is now considered.

The equations in (III.10.1) are uncoupled by adding and subtracting them:

Adding : $M\dfrac{\partial}{\partial y}(\phi + \psi) + \nabla^2 (\phi + \psi) + 1 = 0$ in D

Subtracting: $M\dfrac{\partial}{\partial y}(\phi - \psi) - \nabla^2 (\phi - \psi) + 1 = 0$ in D

or, letting $a = \phi + \psi$ and $b = \phi - \psi$ (III.10.11)

CHAPTER III

$$M\frac{\partial a}{\partial y} + \nabla^2 a + 1 = 0 \text{ in } D, \quad a = 0 \text{ on } \partial D \qquad \text{(III.10.12)}$$

$$M\frac{\partial b}{\partial y} - \nabla^2 b + 1 = 0 \text{ in } D, \quad b = 0 \text{ on } \partial D \qquad \text{(III.10.13)}$$

If we now let

$$a = e^{-Ky} u, \quad b = e^{Ky} v, \text{ where } K = \frac{M}{2} \qquad \text{(III.10.14)}$$

Equation (III.10.12) becomes

$$\nabla^2 u - K^2 u = -e^{Ky} \text{ in } D, \quad u = 0 \text{ on } \partial D \qquad \text{(III.10.15)}$$

and equation (III.10.13) becomes

$$\nabla^2 v - K^2 v = e^{-Ky} \text{ in } D, \quad V = 0 \text{ on } \partial D \qquad \text{(III.10.16)}$$

Equations (III.10.15) and (III.10.16) are then alternative, equivalent representations of the M.H.D. equations. In terms of u and v, the original variables ϕ and ψ are

$$\phi = \tfrac{1}{2}(e^{-Ky} u + e^{Ky} v) \qquad \text{(III.10.17)}$$

$$\psi = \tfrac{1}{2}(e^{-Ky} u - e^{Ky} v) \qquad \text{(III.10.18)}$$

Equations (III.10.15) and (III.10.16) are both of the form

$$\nabla^2 p - K^2 p = f \text{ in } D, \quad p = 0 \text{ on } \partial D \qquad \text{(III.10.19)}$$

Equation (III.10.19) arises from the gradients of the quadratic convex-concave saddle functional

$$L(p,\underline{q}) = \int_D \left\{ -p \operatorname{div} \underline{q} + \tfrac{1}{2} K^2 p^2 - \tfrac{1}{2} \underline{q}^2 + pf \right\} dD \qquad \text{(III.10.20)}$$

as the gradients are:

$$\nabla_p L(p,\underline{q}) = -\operatorname{div} \underline{q} + K^2 p + f \qquad \text{(III.10.21)}$$

$$\nabla_{\underline{q}} L(p,\underline{q}) = \operatorname{grad} p - \underline{q} \qquad \text{(III.10.22)}$$

and setting $\nabla_p L(p,\underline{q}) = \nabla_{\underline{q}} L(p,\underline{q}) = 0$ gives

$\nabla^2 p - K^2 p = f$, as required, noting that $p = 0$ on ∂D.

Comparing equation (III.10.20) with the equation of the general quadratic, equation (III.2.1), it can be seen that we have taken

$$A = -\operatorname{div}, \quad A^{x} = \operatorname{grad}, \quad B = K^2 I, \quad C = I \text{ and } g = 0 \qquad \text{(III.10.23)}$$

Theorem (III.8.1) iterations can be applied to this problem; and, from the work on the convergence of an unbounded operator in the previous section, the iterations will converge to the unique solution of equation (III.10.19), if it exists, provided the following conditions are satisfied:

(a) $-\nabla^2$ and $K^2 I$ are closed, symmetric operators with dense domains in a Hilbert space H such that

$$D(-\nabla^2) \subseteq D(-\nabla^2 + K^2 I)$$

(b) There exists a real number Y such that

$$\int_D -x\,\nabla^2 x\,dD \geqslant Y^2 \int_D x^2\,dD \quad \forall\, x \in D(-\nabla^2)$$

(c) There exists real numbers n and N such that

$$(n-1)\int_D -x\,\nabla^2 x\,dD \leq \int_D K^2 x^2\,dD \leq (N-1)\int_D -x\,\nabla^2 x\,dD$$

$$\forall\, x \in D(-\nabla^2)$$

(a) Using definitions (II.5.11) and (II.5.12) it can be shown that the operators $-\nabla^2$ and $K^2 I$ are closed operators with dense domains; if the domains are taken as the same, $D(-\nabla^2) \subseteq D(-\nabla^2 + K^2 I)$ is satisfied. Finally, by equation (II.5.6), both these operators are symmetric.

(b) As we require all $x \in D(-\nabla^2)$ to vanish on the boundary, then

$$\int_D -x\nabla^2 x\,dD = \int_D (\nabla x)^2\,dD.$$

We therefore require there to exist $Y \in \mathbb{R}$ such that

$$\int_D x^2\,dD \leq \frac{1}{Y^2}\int_D (\nabla x)^2\,dD \quad \forall\, x \in D(-\nabla^2) \qquad \text{(III.10.24)}$$

From page 146 of (44),

$\int_D x^2\,dD \leq \frac{1}{\mu^2}\int_D (\nabla x)^2\,dD$, where $\frac{1}{\mu}$ is the projection of D onto the x-axis. (III.10.25)

If we therefore let $Y = \mu$, then (III.10.24) is satisfied.

(c) Similarly we therefore require there to exist n and $N \in \mathbb{R}$ such that

$$(n-1) \int_D (\nabla x)^2 \, dD \leq K^2 \int_D x^2 \, dD \leq (N-1) \int_D (\nabla x)^2 \, dD$$

$$\forall x \in D(-\nabla^2) \qquad \text{(III.10.26)}$$

The left-hand side of this equation is easily satisfied, as $\int_D (\nabla x)^2 \, dD$ and $\int_D x^2 \, dD$ are both positive quantities, and we can take n such that $n - 1 \leq 0$.

The right hand inequality can be written

$$\int_D x^2 \, dD \leq \frac{(N-1)}{K^2} \int_D (\nabla x)^2 \, dD$$

Now, by equation (III.10.25),

$\int_D x^2 \, dD \leq \frac{1}{\mu^2} \int_D (\nabla x)^2 \, dD$; so if we choose N so that $\frac{N-1}{K^2} = \frac{1}{\mu^2}$, then

$$\int_D x^2 \, dD \leq \frac{1}{\mu^2} \int_D (\nabla x)^2 \, dD = \frac{N-1}{K^2} \int_D (\nabla x)^2 \, dD$$

and hence the right hand inequality is satisfied.

Therefore, the conditions for convergence are all satisfied. Applying theorem (III.8.1) iterations to the problem gives the algorithm:

Choose p_o, $p_o = 0$ on ∂D

For $n = 0,1,2,\ldots$ find p_{2n+1}: $\nabla^2 p_{2n+1} = K^2 p_{2n} + f$,

$$p_{2n+1} = 0 \text{ on } \partial D$$

Find $$\lambda_{2n+2} = \frac{\int_D K^2 (p_{2n} - p_{2n+1})^2 \, dD}{\int_D \{K^2 (p_{2n} - p_{2n+1})^2 + (\nabla (p_{2n} - p_{2n+1}))^2\} \, dD}$$

Let $p_{2n+2} = \lambda_{2n+2} (p_{2n} - p_{2n+1}) + p_{2n+1}$

Then $\lim_{n \to \infty} \| p_{2n+2} - p_e \| = 0$, where p_e is the solution of

$$\nabla^2 p - K^2 p = f \text{ in } D, \quad p = 0 \text{ on } \partial D \qquad \text{(III.10.27)}$$

Applying this iteration with $f = -e^{Ky}$ then $f = e^{-Ky}$ will give u_{2n+2} and v_{2n+2} respectively; then equations (III.10.17) and (III.10.18) can

(c) Similarly we therefore require there to exist n and $N \in \mathbb{R}$ such that

$$(n-1) \int_D (\nabla x)^2 \, dD \leq K^2 \int_D x^2 \, dD \leq (N-1) \int_D (\nabla x)^2 \, dD$$

$$\forall x \in D(-\nabla^2) \qquad \text{(III.10.26)}$$

The left-hand side of this equation is easily satisfied, as $\int_D (\nabla x)^2 \, dD$ and $\int_D x^2 \, dD$ are both positive quantities, and we can take n such that $n - 1 \leq 0$.

The right hand inequality can be written

$$\int_D x^2 \, dD \leq \frac{(N-1)}{K^2} \int_D (\nabla x)^2 \, dD$$

Now, by equation (III.10.25),

$\int_D x^2 \, dD \leq \frac{1}{\mu^2} \int_D (\nabla x)^2 \, dD$; so if we choose N so that $\frac{N-1}{K^2} = \frac{1}{\mu^2}$, then

$$\int_D x^2 \, dD \leq \frac{1}{\mu^2} \int_D (\nabla x)^2 \, dD = \frac{N-1}{K^2} \int_D (\nabla x)^2 \, dD$$

and hence the right hand inequality is satisfied.

Therefore, the conditions for convergence are all satisfied. Applying theorem (III.8.1) iterations to the problem gives the algorithm:

Choose p_o, $p_o = 0$ on ∂D

For $n = 0,1,2,\ldots$ find p_{2n+1}: $\nabla^2 p_{2n+1} = K^2 p_{2n} + f$,

$$p_{2n+1} = 0 \text{ on } \partial D$$

Find $$\lambda_{2n+2} = \frac{\int_D K^2 (p_{2n} - p_{2n+1})^2 \, dD}{\int_D \{K^2 (p_{2n} - p_{2n+1})^2 + (\nabla (p_{2n} - p_{2n+1}))^2\} \, dD}$$

Let $p_{2n+2} = \lambda_{2n+2} (p_{2n} - p_{2n+1}) + p_{2n+1}$

Then $\lim_{n \to \infty} \| p_{2n+2} - p_e \| = 0$, where p_e is the solution of

$$\nabla^2 p - K^2 p = f \text{ in } D, \quad p = 0 \text{ on } \partial D \qquad \text{(III.10.27)}$$

Applying this iteration with $f = -e^{Ky}$ then $f = e^{-Ky}$ will give u_{2n+2} and v_{2n+2} respectively; then equations (III.10.17) and (III.10.18) can be used to find the approximations to the solutions of the original M.H.D. equation, ϕ_{2n+2} and ψ_{2n+2}, and we will have

$$\lim_{n \to \infty} \left\| \phi_{2n+2} - \phi_e \right\| + \left\| \psi_{2n+2} - \psi_e \right\| = 0.$$

CHAPTER III

III.11 Application of Iterative Methods to the problem $\nabla^2\phi = F'(\phi)$, $F(\phi)$ Convex

Let $L(\phi,\underline{u}) = \int_V \left\{ -\phi \operatorname{div} \underline{u} - \frac{1}{2}\underline{u}^2 + F(\phi) \right\} dV$ (III.11.1)

where ϕ belongs to the set of differentiable functions which disappear on the boundary, and $F(\phi)$ is a twice-differentiable function of ϕ such that $F(\phi): X \rightarrow X$ where X is the set of differentiable functions.

The functional derivatives of $L(\phi,\underline{u})$ are:

$$\nabla_\phi L(\phi,\underline{u}) = -\operatorname{div}\underline{u} + F'(\phi) \qquad \text{(III.11.2)}$$

$$\nabla_u L(\phi,\underline{u}) = \operatorname{grad}\phi - \underline{u} \qquad \text{(III.11.3)}$$

(See example 2 after definition (II.7.1) for the derivative of $\int_V F(\phi)\, dV$)

Setting $\nabla_\phi L(\phi,\underline{u}) = \nabla_u L(\phi,\underline{u}) = 0$ gives the problem

$$\nabla^2\phi = F'(\phi) \text{ on } V, \quad \phi = 0 \text{ on } \partial V \qquad \text{(III.11.4)}$$

Using equation (II.9.1), $L(\phi,\underline{u})$ is a convex-concave saddle functional if

$$\int_V \Big\{ -\phi_1 \operatorname{div}\underline{u}_1 - \tfrac{1}{2}\underline{u}_1^2 + F(\phi_1) + \phi_2 \operatorname{div}\underline{u}_2 + \tfrac{1}{2}\underline{u}_2^2 - F(\phi_2)$$
$$- (\phi_1 - \phi_2)(-\operatorname{div}\underline{u}_2 + F'(\phi_2)) - (\underline{u}_1 - \underline{u}_2)(\operatorname{grad}\phi_1 - \underline{u}_1) \Big\}\, dV \geq 0$$

or $$\int_V \left\{ \tfrac{1}{2}(\underline{u}_1 - \underline{u}_2)^2 + F(\phi_1) - F(\phi_2) - (\phi_1 - \phi_2) F'(\phi_2) \right\} dV \geq 0 \qquad \text{(III.11.5)}$$

By equation (II.8.3), equation (III.11.5) is satisfied provided $F(\phi)$ is a convex function; and, as $\int_V \frac{1}{2}(\underline{u}_1 - \underline{u}_2)^2\, dV$ is only equal to zero when $\underline{u}_1 = \underline{u}_2$, $L(\phi,\underline{u})$ is a strict convex-concave saddle functional provided $F(\phi)$ is convex for all ϕ.

The rest of this section applies three different methods to the problem: classical dual extremum principles, cobweb iteration and optimising iteration (which can only be used for linear $F'(\phi)$). The methods are then applied to a simple example, and the results obtained are compared.

CHAPTER III

(a) Classical Dual Extremum Principles

Using theorem (II.12.1), the classical dual extremum principles are:

$$L(\phi_\beta, \underline{u}_\beta) \le L(\phi_e, \underline{u}_e) \le L(\phi_\alpha, \underline{u}_\alpha) \qquad (III.11.6)$$

where

$$L(\phi_\beta, \underline{u}_\beta) = \int_V \left\{ F(\phi_\beta) - \phi_\beta F'(\phi_\beta) - \tfrac{1}{2} \underline{u}_\beta^{\,2} \right\} dV \qquad (III.11.7)$$

$$\text{and } -\operatorname{div} \underline{u}_\beta + F'(\phi_\beta) = 0 \text{ on } V, \quad \phi_\beta = 0 \text{ on } \partial V \qquad (III.11.8)$$

$$L(\phi_\alpha, \underline{u}_\alpha) = \int_V \left\{ \tfrac{1}{2} (\operatorname{grad} \phi_\alpha)^2 + F(\phi_\alpha) \right\} dV \qquad (III.11.9)$$

$$\text{and } \phi_\alpha = 0 \text{ on } \partial V \qquad (III.11.10)$$

$$\text{and } L(\phi_e, \underline{u}_e) = \int_V \left\{ \tfrac{1}{2} (\operatorname{grad} \phi_e)^2 + F(\phi_e) \right\} dV \qquad (III.11.11)$$

$$\text{where } \nabla^2 \phi_e = F'(\phi_e) \text{ in } V, \quad \phi_e = 0 \text{ on } \partial V \qquad (III.11.12)$$

The method consists of finding functions $(\phi_\beta, \underline{u}_\beta)$ which satisfy equation (III.11.8) and ϕ_α which satisfies equation (III.11.10), and then computing $L(\phi_\beta, \underline{u}_\beta)$ and $L(\phi_\alpha, \underline{u}_\alpha)$; it is hoped that the smaller is $L(\phi_\alpha, \underline{u}_\alpha) - L(\phi_\beta, \underline{u}_\beta)$, the nearer to ϕ_e are ϕ_α and ϕ_β (although this is not always true, as we discussed earlier in section III.3). The method is essentially trial and error, although trial functions can be chosen which contain parameters to be optimised. The method can be applied to problems with non-linear $F'(\phi)$. (REFERENCE: (59)).

(b) Cobweb iteration

For the functional given in equation (III.11.1), cobweb iterative scheme C from section III.3 becomes

Choose $(\phi_0, \underline{u}_0)$: $\nabla_{\underline{u}} L(\phi_0, \underline{u}_0) = 0$: $\operatorname{grad} \phi_0 - \underline{u}_0 = 0$ in V

$\phi_0 = 0$ on ∂V

Then, for n = 0,1,2,... Find $\underline{u}_{n+1}$: $\nabla_\phi L(\phi_n, \underline{u}_{n+1}) = 0$:

$-\operatorname{div} \underline{u}_{n+1} + F'(\phi_n) = 0$ in V

Find ϕ_{n+1}: $\nabla_{\underline{u}} L(\phi_{n+1}, \underline{u}_{n+1}) = 0$:

$\operatorname{grad} \phi_{n+1} - \underline{u}_{n+1} = 0$ in V,

$\phi_{n+1} = 0$ on ∂V (III.11.13)

Then

$$L(\phi_n, \underline{u}_{n+1})_\beta = \int_V \left\{ - \phi_n F'(\phi_n) - \tfrac{1}{2} \underline{u}_{n+1}^2 + F(\phi_n) \right\} dV \quad \text{(III.11.14)}$$

and $$L(\phi_n, \underline{u}_n)_\alpha = \int_V \left\{ \tfrac{1}{2} (\nabla \phi_n)^2 + F(\phi_n) \right\} dV \quad \text{(III.11.15)}$$

$$L(\phi_n, \underline{u}_n)_\alpha - L(\phi_n, \underline{u}_{n+1})_\beta = \tfrac{1}{2} \int_V (\underline{u}_n - \underline{u}_{n+1})^2 \, dV \quad \text{(III.11.16)}$$

It has not been possible to prove that

$$\lim_{n \to \infty} \left\{ L(\phi_n, \underline{u}_n)_\alpha - L(\phi_n, \underline{u}_{n+1})_\beta \right\} = 0$$

although in particular examples convergence could occur. This method can also be used for problems with non-linear $F'(\phi)$.

(c) Optimising Iteration

This method can only be used for problems in which $F'(\phi)$ is linear; this is due to the requirement, from section III.5, that whenever

$$\nabla_\phi L(\phi_{2n}, \underline{u}_{2n}) = 0 \text{ and } \nabla_\phi L(\phi_{2n+1}, \underline{u}_{2n+1}) = 0 \text{ then}$$

$$\nabla_\phi L(\lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}, \lambda_{2n+2} (\underline{u}_{2n} - \underline{u}_{2n+1}) + \underline{u}_{2n+1}) = 0.$$

Using equation (III.11.2) we require that whenever

$$- \operatorname{div} \underline{u}_{2n} + F'(\phi_{2n}) = 0 \text{ and } - \operatorname{div} \underline{u}_{2n+1} + F'(\phi_{2n+1}) = 0$$

then

$$- \operatorname{div} (\lambda_{2n+2} (\underline{u}_{2n} - \underline{u}_{2n+1}) + \underline{u}_{2n+1}) + F'(\lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}) = 0$$

and this is only true if $F'(\phi)$ is linear. For this reason, therefore, we will consider the application of the optimising iteration to the problem with

$F(\phi) = \frac{1}{2} K^2 \phi^2 + b$, $(K, b \in \mathbb{R})$. This results in the problem

$$\nabla^2 \phi = K^2 \phi + b \text{ in } V, \quad \phi = 0 \text{ on } \partial V \quad \text{(III.11.17)}$$

the functional given by equation (III.11.1) becomes

$$L(\phi, \underline{u}) = \int_V \left\{ - \phi \operatorname{div} \underline{u} - \tfrac{1}{2} \underline{u}^2 + \tfrac{1}{2} K^2 \phi^2 + b\phi \right\} dV \quad \text{(III.11.18)}$$

and the functional derivatives are then

$$\nabla_\phi L(\phi, \underline{u}) = - \operatorname{div} \underline{u} + K^2 \phi + b \quad \text{(III.11.19)}$$

$$\nabla_{\underline{u}} L(\phi, \underline{u}) = \operatorname{grad} \phi - \underline{u} \quad \text{(III.11.20)}$$

CHAPTER III

It was shown in section III.10 that if theorem (III.8.1) iterations are applied to the problem, the iterations will converge to the unique solution of equation (III.11.17).

The rest of this section applies the three methods given above to the same simple example. The purpose of this is to show the advantage of the iterative methods, particularly the optimising iterative method, in terms of the work involved and the accuracy obtained.

Consider the problem

$$\nabla^2\phi = \phi + 1 \text{ on the unit disc } 0 \le r \le 1, \quad 0 \le \theta \le 2\pi \qquad \text{(III.11.21)}$$

with $\phi(1,\theta) = 0$ and $\phi(0,\theta)$ bounded.

Also assume that ϕ is a function of r only; then the inner product is given by

$$\int_V pq \, dV = 2\pi \int_0^1 pq \, r \, dr \qquad \text{(III.11.22)}$$

$$F'(\phi) = \phi + 1, \text{ so } F(\phi) = \tfrac{1}{2}\phi^2 + \phi \qquad \text{(III.11.23)}$$

In polar co-ordinates, (22), p 242

$$\nabla^2\phi = \frac{1}{r}\frac{\partial}{\partial r}\left(r \frac{\partial\phi}{\partial r}\right) \qquad \text{(III.11.24)}$$

$$\text{grad}\,\phi = \frac{\partial\phi}{\partial r}\,\underline{e}r \qquad \text{(III.11.25)}$$

$$\text{div}\,\underline{u} = \frac{1}{r}\frac{d}{\partial r}\left(r\underline{u}\right) \quad \text{where } \underline{u} = \underline{u}\,er \qquad \text{(III.11.26)}$$

(a) <u>Classical Dual Extremum Principles</u>

From equations (III.11.7) - (III.11.12), and (III.11.22), we have

$$L(\phi_e, u_e) = 2\pi \int_0^1 \left\{ \tfrac{1}{2}(\text{grad}\phi_e)^2 + \tfrac{1}{2}\phi_e^2 + \phi_e \right\} r \, dr \qquad \text{(III.11.27)}$$

$$\text{where } \nabla^2\phi_e = \phi_e + 1, \quad \phi_e(1,\theta) = 0 \qquad \text{(III.11.28)}$$

$$\phi(0,\theta) \text{ is bounded}$$

$$L(\phi_\beta, u_\beta) = -\pi \int_0^1 \left\{ \phi_\beta^2 + u_\beta^2 \right\} r\,dr \qquad \text{(III.11.29)}$$

$$\text{where } \frac{1}{r}\frac{\partial}{\partial r}(r u_\beta) = \phi_\beta + 1 \text{ in V}, \quad \phi_\beta(0,\theta) \text{ is bounded and } \phi_\beta(1,\theta) = 0 \qquad \text{(III.11.30)}$$

$$L(\phi_\alpha, u_\alpha) = 2\pi \int_0^1 \left\{ \tfrac{1}{2} \left(\frac{\partial \phi}{\partial r}\right)^2 + \tfrac{1}{2} \phi_\alpha^2 + \phi_\alpha \right\} r\,dr \qquad \text{(III.11.31)}$$

$$\text{where } \phi_\alpha(1,\theta) = 0, \quad \phi(0,\theta) \text{ is bounded} \qquad \text{(III.11.32)}$$

The trial function ϕ which we are going to use to illustrate the method is $\phi = m(1 - r^n)$, where m and n are parameters to be determined.

Lower Bound

$$\text{Let } \phi = m(1 - r^n), \quad u_\beta = m\left(\frac{r}{2} - \frac{r^{n+1}}{n+2}\right) + \frac{r}{2} \qquad \text{(III.11.33)}$$

Then equation (III.11.30) is satisfied. Using equation (III.11.29),

$$L(\phi_\beta, u_\beta) = -\pi \int_0^1 \left\{ m^2 (1 - r^n)^2 + \left(m\left(\frac{r}{2} - \frac{r^{n+1}}{n+2}\right) + \frac{r}{2}\right)^2 \right\} r\,dr$$

$$= -\frac{\pi}{16}\left[m^2 \left(\frac{9n^5 + 75n^4 + 190n^3 + 148n^2 + 32n + 28}{(n+1)(n+4)(n+2)^3}\right) + m\left(\frac{2n(n+6)}{(n+2)(n+4)}\right) + 1 \right] \qquad \text{(III.11.34)}$$

Equation (III.11.34) is first optimised with respect to m and then with respect to n; as $L(\phi_\beta, u_\beta)$ is the lower bound, we need to find local maximum of $L(\phi_\beta, u_\beta)$.

$$\frac{\partial L(\phi_\beta, u_\beta)}{\partial m} = -\frac{\pi}{8}\left[m\left(\frac{9n^5 + 75n^4 + 190n^3 + 148n^2 + 32n + 28}{(n+1)(n+4)(n+2)^3}\right) + \left(\frac{n(n+6)}{(n+2)(n+4)}\right)\right]$$

Local extrema occur when $\frac{\partial L(\phi_\beta, u_\beta)}{\partial m} = 0$, that is

$$m = \frac{-n(n+1)(n+6)(n+2)^3}{(9n^5 + 75n^4 + 190n^3 + 148n^2 + 32n + 28)} \qquad \text{(III.11.35)}$$

Substituting equation (III.11.35) into (III.11.34) gives

$$L(\phi_\beta, u_\beta) = \frac{\pi}{16}\left[\frac{n^2 (n+1)(n+2)(n+6)^2}{(n+4)(9n^5 + 75n^4 + 190n^3 + 148n^2 + 32n + 28)}\right]^{-1} \qquad \text{(III.11.36)}$$

$L(\phi_\beta, u_\beta)$ given by equation (III.11.36) will have local maxima at values of n for which $\frac{\partial^2 L(\phi_\beta, u_\beta)}{\partial m^2}$ is negative; that is

$$\text{when } \frac{\partial^2 L(\phi_\beta, u_\beta)}{\partial m^2} = -\frac{\pi}{8}\left[\frac{9n^5 + 75n^4 + 190n^3 + 148n^2 + 32n + 28}{(n+1)(n+4)(n+2)^3}\right] \leq 0 \qquad \text{(III.11.37)}$$

We therefore need to find the values of n which maximises equation (III.11.36), and then check to see for which n equation (III.11.37) is satisfied.

$$\frac{\partial L(\phi_\beta, u_\beta)}{\partial n} = \frac{\pi}{4}\left[\frac{(-6n^{10} - 88n^9 - 426n^8 - 540n^7 + 1849n^6 + 7956n^5 + 15006n^4 + 18584n^3 + 13896n^2 + 4032n)}{((n+4)(9n^5 + 75n^4 + 190n^3 + 148n^2 + 32n + 28))^2}\right] \qquad \text{(III.11.38)}$$

$$\text{Let } \frac{\partial L(\phi_\beta, u_\beta)}{\partial n} = \frac{\pi}{4}\,\frac{f(n)}{(g(n))^2} \qquad \text{(III.11.39)}$$

$$\text{Then } \frac{\partial^2 L(\phi_\beta, u_\beta)}{\partial n^2} = \frac{\pi}{4}\left[\frac{(g(n)^2 f'(n) - 2 f(n) g(n) g'(n))}{(g(n))^4}\right] \qquad \text{(III.11.40)}$$

Local maxima of $L(\phi_\beta, u_\beta)$ occur when $f(n) = 0$ and $\frac{\partial^2 L(\phi_\beta, u_\beta)}{\partial n^2} \leq 0$ - that is, when $f'(n) \leq 0$.

Hence the maxima of $L(\phi_\beta, u_\beta)$ occur at values of n for which

(i) $f(n) = -6n^{10} - 88n^9 - 426n^8 - 540n^7 + 1849n^6 + 7956n^5 + 15006n^4 + 18584n^3 + 13896n^2 + 4032n = 0$

For these values of n we require

(ii) $f'(n) = (-60n^9 - 792n^8 - 3408n^7 - 3780n^6 + 11094n^5 + 39780n^4 + 60024n^3 + 55752n^2 + 27792n + 4032) \leq 0$

and

(iii), from equation (III.11.37), we require

$$h(n) = \frac{(9n^5 + 75n^4 + 190n^3 + 148n^2 + 32n + 28)}{(n+1)(n+4)(n+2)^3} \geq 0$$

Suitable values of n are then substituted into equation (III.11.36) to find the maximum $L(\phi_\beta, u_\beta)$; the n which gives this maximum is then substituted into equation (III.11.35) to find the best value of m.

Results

Roots of f(n)	Sign of $f'(n)$	Sign of h(n)	$L(\phi_\beta, u_\beta)$ for those n for which $f'(n) \leq 0$, $h(n) \geq 0$
0	+	+	-
-6	+	-	-
-4.109897884	-	+	-13.95800594
-3.278465654	+	-	-
-0.607174869	-	+	- 0.185053213
2.859952296	-	+	- 0.168801019

TABLE III.11.1

Therefore the maximum $L(\phi_\beta, u_\beta)$ is -0.168801019 which occurs when

n = 2.859952296 and m = -0.897146462

$$\phi_\beta = -0.897146462 \ (1 - r^{2.859952296})$$

Upper Bound

Again, we let $\phi_\alpha = m(1 - r^n)$ (III.11.41)

$$\text{Then } L(\phi_\alpha, u_\alpha) = 2\pi \int_0^1 \left\{ \frac{m^2}{2} (1 - r^n)^2 + n^2 r^2 (n-1) + m(1 - r^n) \right\} r \, dr$$

$$= \pi \left[\frac{m^2}{2} \frac{n(n^2 + 4n + 2)}{(n+1)(n+2)} + m \frac{n}{n+2} \right] \qquad \text{(III.11.42)}$$

As in the lower bound, $L(\phi_\alpha, \underline{u}_\alpha)$ is first optimised with respect to m and then n.

$$\frac{\partial L(\phi_\alpha, \underline{u}_\alpha)}{\partial m} = \pi \left[m \left(\frac{n\,(n^2 + 4n + 2)}{(n+1)\,(n+2)} \right) + \frac{n}{n+2} \right]$$

Local extremum occur when $\frac{\partial L(\phi_\alpha, \underline{u}_\alpha)}{\partial m} = 0$, giving

$$m = \frac{-(n+1)}{(n^2 + 4n + 2)} \qquad \text{(III.11.43)}$$

Substituting equation (III.11.43) into (III.11.42) gives

$$L(\phi_\alpha, \underline{u}_\alpha) = -\frac{\pi}{2}\,\frac{n\,(n+1)}{(n+2)\,(n^2 + 4n + 2)} \qquad \text{(III.11.44)}$$

Local minima will occur when $\frac{\partial^2 L(\phi_\alpha, \underline{u}_\alpha)}{\partial m^2} \geqslant 0$, that is when

$$h(n) = \frac{n\,(n^2 + 4n + 2)}{(n+1)\,(n+2)} \geqslant 0 \qquad \text{(III.11.45)}$$

Using equation (III.11.44),

$$\frac{\partial L(\phi_\alpha, \underline{u}_\alpha)}{\partial n} = \frac{\pi}{2} \left[\frac{(n^4 + 2n^3 - 4n^2 - 8n - 4)}{(n+2)^2\,(n^2 + 4n + 2)^2} \right] \qquad \text{(III.11.46)}$$

$$\text{Let } \frac{\partial L(\phi_\alpha, \underline{u}_\alpha)}{\partial n} = \frac{\pi}{2}\,\frac{f(n)}{(g(n))^2} \qquad \text{(III.11.47)}$$

$$\text{Then } \frac{\partial^2 L(\phi_\alpha, \underline{u}_\alpha)}{\partial n^2} = \frac{\pi}{2} \left[\frac{(g\,(n)^2\,f'\,(n) - 2f\,(n)\,g\,(n)\,g'\,(n)}{(g\,(n)\,)^2} \right] \qquad \text{(III.11.48)}$$

Using the same analysis as in the lower bound, the minima of $L(\phi_\alpha, \underline{u}_\alpha)$ occur at values of n for which

(i) $\quad f\,(n) = n^4 + 2n^3 - 4n^2 - 8n - 4 = 0$

For these values of n we require

(ii) $\quad f'(n) = 4n^3 + 6n^2 - 8n - 8 \geqslant 0$

and

(iii) $\quad h(n) = \frac{n\,(n^2 + 4n + 2)}{(n+1)\,(n+2)} \geqslant 0$

The minimum of $L(\phi_\alpha, \underline{u}_\alpha)$ is then found by substituting suitable n into equation (III.11.44) and the best m is found by using equation (III.11.43).

Results

Roots of $f(n)$	Sign of $f'(n)$	Sign of $h(n)$	$L(\phi_\alpha, u_\alpha)$ for those n for which $f'(n) \geq 0$ and $h(n) \geq 0$
-2.581545954	-	+	-
2.112009744	+	+	-0.168408901

TABLE III.11.2

The minimum $L(\phi_\alpha, u_\alpha)$ is -0.168408901 which occurs when

$n = 2.112009744$ and $m = -0.208738896$:

$$= -0.208738896\ (1 - r^{2.112009744}).$$

We now have

$-0.168801019 \leq L(\phi_e, u_e) \leq -0.168408901$ and the upper and lower bounds are close; but the method does not indicate how close we are to ϕ_e, as we do not know if ϕ_e is of the form $m(1 - r^n)$.

This method involved a large amount of calculation which needed computing facilities; obviously a more complicated trial function ϕ would require a great deal more calculations.

(b) Cobweb Iteration

From equations (III.11.13) - (III.11.15), and (III.11.22) - (III.11.20) the cobweb iterative scheme for the problem is

Choose (ϕ_o, u_o) : $\frac{\partial \phi_o}{\partial r}\, \underline{e}_r = \underline{u}_o$ in V, $\phi_o\,(1,\theta) = 0$

$\phi_o\,(0,\theta)$ is bounded

Then for $n = 0,1,2,\ldots$ Find $\underline{u}_{n+1}$: $\frac{1}{r}\frac{\partial}{\partial r}\left(r\, u_{n+1}\right) = \phi_{n+1}$ in V

where $\underline{u}_{n+1} = u_{n+1}\, \underline{e}_r$

Find ϕ_{n+1} : $\frac{\partial \phi_{n+1}}{\partial r}\, \underline{e}_r = \underline{u}_{n+1}$ in V,

$\phi_{n+1}\,(1,\theta) = 0$

$\phi_{n+1}\,(0,\theta)$ is bounded

(III.11.49)

$$L(\phi_n, u_n + 1)_\beta = -\pi \int_0^1 \left\{ \phi_n^2 + \underline{u}_n^2 + 1 \right\} r\, dr$$

$$L(\phi_n, u_n)_\alpha = \pi \int_0^1 \left\{ \phi_n^2 + \left(\frac{\partial \phi_n}{\partial r} \right)^2 + 2\phi_n \right\} r\, dr \qquad \text{(III.11.49)}$$

Brief details of an iteration starting with $\phi_o = 0$ are given in Appendix V and the results are tabulated below.

n	u_n	ϕ_n	$L(\phi_n, u_{n+1})_\beta$	$L(\phi_n, u_n)_\alpha$
0		0		0
1	$\frac{r}{2}\, \underline{e}_r$	$-\frac{1}{4}(1 - r^2)$	-0.19634954	-0.163624617
2	$\frac{1}{16}(6r + r^3)\, \underline{e}_r$	$-\frac{1}{64}(13 - 12r - r^4)$	-0.169249213	-0.168277692
3	$\frac{1}{384}(153r + 18r^2 + r^5)\, \underline{e}_r$	$-\frac{1}{2304}(487 - 459r^2 - 27r^4 - r^6)$	-0.168445649	-0.168416608

TABLE III.11.3

The iteration seems to be converging to a solution of the form $\sum_{n=0}^{\infty} a_{2n}\, r^{2n}$; a comparison of the accuracy obtained and the work required will be made after the optimising iteration has been considered.

(c) Optimising Iteration

For this problem, theorem (III.8.1) iterations become:

Choose $(\phi_o, u_o) : \frac{\partial \phi_o}{\partial r}\, \underline{e}_r - u_o = 0$, $\phi_o(1,\theta) = 0$
$\phi_o(0,\theta)$ is bounded

Then, for n = 0,1,2,3,...

Find u_{2n+1}: $\frac{1}{r}\frac{\partial}{\partial r}\left(r\, \underline{u}_{2n+1} \right) = \phi_{2n+1}$, $\underline{u}_{2n+1} = u_{2n+1}\, \underline{e}_r$

Find ϕ_{2n+1}: $\frac{\partial \phi_{2n+1}}{\partial r}\, \underline{e}_r = \underline{u}_{2n+1}$, $\phi_{2n+1}(1,\theta) = 0$
$\phi_{2n+1}(0,\theta)$ is bounded

Find $\lambda_{2n+2} = \dfrac{\int_0^1 (\phi_{2n} - \phi_{2n+1})^2 \, r\,dr}{\int_0^1 \left\{ (\phi_{2n} - \phi_{2n+1})^2 + \frac{\partial}{\partial r}(\phi_{2n} - \phi_{2n+1})^2 \right\} r\,dr}$

Let $\phi_{2n+2} = \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$

$\underline{u}_{2n+2} = \lambda_{2n+2}(\underline{u}_{2n} - \underline{u}_{2n+1}) + \underline{u}_{2n+1}$

$$L(\phi_{2n}, \underline{u}_{2n})_\alpha = \pi \int_0^1 \left\{ \phi_{2n}^2 + \left(\frac{\partial \phi_{2n}}{\partial r}\right)^2 + 2\phi_{2n} \right\} r\,dr$$

$$L(\phi_{2n}, \underline{u}_{2n+1})_\beta = -\pi \int_0^1 \left\{ \phi_{2n}^2 + \underline{u}^2_{2n+1} \right\} r\,dr \qquad \text{(III.11.50)}$$

Brief details of an iteration starting with $\phi_0 = 0$ are given in Appendix V and the results are given in Table (III.11.4). Only the ϕ_{2n}, $\underline{u}_{2n+1}$, $L(\phi_{2n}, \underline{u}_{2n})_\alpha$ and $L(\phi_{2n}, \underline{u}_{2n+1})_\beta$ have been given in this table; other iterates are in Appendix V.

n	$\underline{u}_{2n+1}$	ϕ_{2n}	$L(\phi_{2n}, \underline{u}_{2n+1})_\beta$	$L(\phi_{2n}, \underline{u}_{2n})_\alpha$
0	$\frac{r}{2}\,\underline{e}_r$	0	-0.19634954	0
1	$\frac{1}{56}(22r + 3r^3)\,\underline{e}_r$	$\frac{-3}{14}(1 - r^2)$	-0.168424829	-0.168299606
2	$\frac{1}{6944}(2743r + 342r^3 + 15r^5)\,\underline{e}_r$	$\frac{-1}{3472}(729 - 684r^2 - 45r^4)$	-0.168420903	-0.16842078
3		$\frac{1}{23484608}(-4935366 + 4636866r^2 + 291150r^4 + 7350r^6)$		-0.16842088

TABLE (III.11.4)

As we showed earlier in this section that the optimising iteration converges to the unique solution of equation (III.11.21) we can now be certain that the solution is of the form $\sum_{n=0}^{\infty} a_{2n} r^{2n}$. It is obvious that both the cobweb and optimising iterations give more accurate approximations to both ϕ_e and $L(\phi_e, u_e)$, with less work. We now need to compare the two iterative methods: for this we need to find the solution ϕ_e and $L(\phi_e, u_e)$.

A particular solution of $\nabla^2\phi = \phi + 1$ is $\phi = -1$;

therefore let $\phi = -1 + \sum_{n=0}^{\infty} a_{2n} r^{2n}$ (III.11.51)

$$= -1 + a_0 + a_2 r^2 + a_4 r^4 + \ldots\ldots \quad \text{(III.11.52)}$$

$$\nabla^2\phi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right) = \sum_{n=1}^{\infty} (2n)^2 a_{2n} r^{2n-2}$$

$$= 2^2 a_2 + 4^2 a_4 r^2 + 6^2 a_6 r^4 + \ldots\ldots \quad \text{(III.11.53)}$$

Then

$$\nabla^2\phi - \phi - 1 = 0 = (2^2 a_2 - a_0) + (4^2 a_4 - a_2) r^2 + (6^2 a_6 - a_4) + \ldots$$

which is satisfied if $2^2 a_2 - a_0 = 0$

$$4^2 a_4 - a_2 = 0$$

$$6^2 a_6 - a_4 = 0$$

that is, $a_{2n} = \dfrac{a_0}{2^{2n}(n!)^2}$, $n = 1, 2, 3, \ldots.$ (III.11.54)

Therefore $\phi = -1 + a_0 \sum_{n=0}^{\infty} \dfrac{r^{2n}}{2^{2n}(n!)^2}$ (III.11.55)

$$\phi(1) = -1 + a_0 \sum_{n=0}^{\infty} \frac{1}{2^{2n}(n!)} = 0$$

giving $a_0 = \dfrac{1}{\sum_{n=0}^{\infty} \dfrac{1}{2^{2n}(n!)}}$

Hence $\phi_e = -1 + \dfrac{\sum_{n=0}^{\infty} \dfrac{r^{2n}}{2^{2n}(n!)^2}}{\sum_{n=0}^{\infty} \dfrac{1}{2^{2n}}(n!)^2}$ (III.11.56)

or, from page 961 of (40),

$$\phi_e = -1 + \frac{I_0(r)}{I_0(1)} \quad \text{(III.11.57)}$$

where $I_0(r)$ is a modified Bessel function.

Then

$L(\phi_e, u_e) = -0.168420889$

(This is derived in Appendix V)

and the first four terms of ϕ_e are

$$\phi_e \simeq -0.21015168 + 0.197462078r^2 + 0.012341379r^4 + 0.0003428161086r^6$$

After three cycles, the cobweb iteration gave

$$\phi_e \simeq -0.211371527 + 0.19921875r^2 + 0.01171875r^4 + 0.0004340277r^6$$

and $-0.168445649 \leq L(\phi_e, u_e) \leq -0.168416608$

and the optimising iteration gave

$$\phi_e \simeq -0.21015322 + 0.197442767r^2 + 0.012397481r^4 + 0.0003129709468r^6$$

and $-0.168420903 \leq L(\phi_e, u_e) \leq -0.16842088$

It is obvious from the results that the optimising iteration gives approximates to ϕ_e and $L(\phi_e, u_e)$ which are a lot more accurate than the approximates given by the cobweb iteration. The work involved in the optimising iterative method is about double that of the work involved in the cobweb iterative method; but this disadvantage is outweighed by the extra accuracy obtained.

It is interesting to note that the lower bounds $L(\phi_{2n}, u_{2n+1})_G$ were increasing for the problem under consideration, even though the method gives no indication as to whether the bounds on the side not being optimised form a monotone sequence. When the method is applied to problems for which convergence cannot be proven, it may not be necessary to carry out two separate iterations giving decreasing upper bounds and increasing lower bounds if one of the iterations give monotone upper and lower bound sequences.

Unless convergence has been proven, it cannot be assumed that the iterates give increasingly better approximations to ϕ_e; however, if the optimising iterative seems to be indicating a solution to the problem of a particular form, it would be worth substituting the suspected solution into the problem to see if the exact solution can be found.

CHAPTER IV

IV.1 Introduction

In this chapter we are interested in obtaining dual extremum principles which offer an improvement over the classical extremum principles given in section II.12 by assuming that the saddle functional $L(\phi, \psi)$ can be decomposed into two saddle functionals $M(\phi, \psi)$ and $N(\phi, \psi)$, with $L(\phi, \psi) = M(\phi, \psi) + N(\phi, \psi)$.

The decomposition dual extremum principles differ from the classical in that both the upper and lower bounds and the functional derivatives each contain an extra trial function which increases the flexibility when choosing trial functions which, in turn can lead to improved upper and lower bounds compared to those obtained in the dual extremum principles.

Section IV.2 sets out the basic theory; the section starts with saddle functionals $M(\phi, \psi)$ and $N(\phi, \psi)$ which are convex/concave; after several lemmas, the theorems on which the rest of this chapter is based are proven.

Section IV.3 applies the main theorem of the chapter to the usual quadratic functional, with

$M(\phi, \psi) = \langle\phi, A_M\psi\rangle + \frac{1}{2}\langle\phi, B_M\phi\rangle - \frac{1}{2}\langle\psi, C_M\psi\rangle + \langle\phi, f_M\rangle + \langle\psi, g_M\rangle$

and $N(\phi, \psi) = \langle\phi, A_N\psi\rangle + \frac{1}{2}\langle\phi, B_N\phi\rangle - \frac{1}{2}\langle\psi, C_N\psi\rangle + \langle\phi, f_N\rangle + \langle\psi, g_N\rangle$

After showing that taking the decomposition operators as multiples of the original operators B and C does not lead to any improvement in the bounds, conditions are derived which ensure that the decomposition bounds do offer an improvement over the classical bounds. The section ends with an example.

Sections IV.4 and IV.5 consider the combination of iterative methods and decomposition dual extremum principles. Section IV.4 develops convergence conditions for the four cobweb iterative schemes introduced in section III.3, and section IV.5 looks at the application of iterative methods to the decomposition dual extremum principles developed in section IV.3. This section ends with an example which compares the conditions necessary for convergence when both cobweb and optimising iterative methods are applied to both classical and decomposition dual extremum principles.

The last section, IV.6, deals with three applications of the decomposition of functionals method.

IV.2 Decomposition dual extremum principles

Let $L(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ be a real functional where E^1 is a subspace of E and F^1 is a subspace of F, and E and F are two inner product spaces. We suppose there exist real convex/concave saddle functionals $M(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ and $N(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ such that there exists a non-trivial decomposition

$$L(\phi, \psi) = M(\phi, \psi) + N(\phi, \psi) \qquad \text{(IV.2.1)}$$

Also suppose that $M(\phi, \psi)$ and $N(\phi, \psi)$ have functional gradients $\nabla_\phi M(\phi, \psi) : E^1 \times F^1 \to E$, $\nabla_\psi M(\phi, \psi) : E^1 \times F^1 \to F$, $\nabla_\phi N(\phi, \psi) : E^1 \times F^1 \to E$ and $\nabla_\psi N(\phi, \psi) : E^1 \times F^1 \to F$, which are found using the definitions in section II.7; then we can prove the lemma which follows from the linearity of the gradients.

Lemma IV.2.1

(a) If $\nabla_\phi M(\phi, \psi) : E^1 \times F^1 \to E$ and $\nabla_\phi N(\phi, \psi) : E^1 \times F^1 \to E$ exist then so does $\nabla_\phi L(\phi, \psi) : E^1 \times F^1 \to E$ and

$$\nabla_\phi L(\phi, \psi) = \nabla_\phi M(\phi, \psi) + \nabla_\phi N(\phi, \psi) \qquad \text{(IV.2.2)}$$
$$\forall \quad (\phi, \psi) \in E^1 \times F^1$$

(b) If $\nabla_\psi M(\phi, \psi) : E^1 \times F^1 \to F$ and $\nabla_\psi N(\phi, \psi) : E^1 \times F^1 \to F$ exist then so does $\nabla_\psi L(\phi, \psi) : E^1 \times F^1 \to F$ and

$$\nabla_\psi L(\phi, \psi) = \nabla_\psi M(\phi, \psi) + \nabla_\psi N(\phi, \psi) \qquad \text{(IV.2.3)}$$
$$\forall \quad (\phi, \psi) \in E^1 \times F^1$$

Proof

(a) Using equation (II.7.8), as the gradients exist,

$$\left[\frac{d}{dt} M(\phi + th, \psi)\right]_{t=0} = \langle h, \nabla_\phi M(\phi, \psi)\rangle \qquad \text{(IV.2.4)}$$

and

$$\left[\frac{d}{dt} N(\phi + th, \psi)\right]_{t=0} = \langle h, \nabla_\phi N(\phi, \psi)\rangle \qquad \text{(IV.2.5)}$$

Then, as $\frac{d}{dt}$ is a linear operator,

$$\left[\frac{d}{dt}\left\{M(\phi + th, \psi) + N(\phi + th, \psi)\right\}\right]_{t=0} = \langle h, \nabla_\phi M(\phi, \psi) + \nabla_\phi N(\phi, \psi)\rangle \qquad \text{(IV.2.6)}$$

From equation (IV.2.1), equation (IV.2.6) implies that

$$\left[\frac{d}{dt} L(\phi + th, \psi)\right]_{t=0} = \langle h, \nabla_\phi M(\phi, \psi) + \nabla_\phi N(\phi, \psi)\rangle \qquad \text{(IV.2.7)}$$

As the right hand side of equation (IV.2.6) exists, so does the left hand side, and

$$\left[\frac{d}{dt} L(\phi + th, \psi)\right]_{t=0} = \langle h, \nabla_\phi L(\phi, \psi)\rangle \qquad \text{(IV.2.8)}$$

by equation (II.7.8). Combining equations (IV.2.7) and (IV.2.8) gives equation (IV.2.2), thus proving the lemma.

(b) Similarly using equation (II.7.9),

$$\left[\frac{d}{dt} M(\phi, \psi + tk)\right]_{t=0} + \left[\frac{d}{dt} N(\phi, \psi + tk)\right]_{t=0}$$

$$= \langle k, \nabla_\psi M(\phi, \psi) + \nabla_\psi N(\phi, \psi)\rangle$$

or

$$\left[\frac{d}{dt}\left\{M(\phi, \psi + tk) + N(\phi, \psi + tk)\right\}\right]_{t=0}$$

$$= \left[\frac{d}{dt} L(\phi, \psi + tk)\right]_{t=0} = \langle k, \nabla_\psi M(\phi, \psi) + \nabla_\psi N(\phi, \psi)\rangle \qquad \text{(IV.2.9)}$$

Then $\left[\frac{d}{dt} L(\phi, \psi + tk)\right]_{t=0}$ exists, and

$$\left[\frac{d}{dt} L(\phi, \psi + tk)\right]_{t=0} = \langle k, \nabla_\psi L(\phi, \psi)\rangle \qquad \text{(IV.2.10)}$$

Combining equations (IV.2.9) and (IV.2.10) thus proves the lemma.

Using the conditions that $M(\phi, \psi)$ and $N(\phi, \psi)$ are convex/concave saddle functionals, and the above lemma, we can prove the following:

Lemma IV.2.2

If $M(\phi, \psi) : E^1 \times F^1 \rightarrow \mathbb{R}$ and $N(\phi, \psi) : E^1 \times F^1 \rightarrow \mathbb{R}$ are convex/concave saddle functionals and $L(\phi, \psi) : E^1 \times F^1 \rightarrow \mathbb{R}$ is defined by $L(\phi, \psi) = M(\phi, \psi) + N(\phi, \psi)$, then $L(\phi, \psi)$ is also a convex/concave saddle functional.

Proof

By theorem (II.9.1), $L(\phi, \psi)$ will be a convex/concave saddle functional if

$$L(\phi_1, \psi_1) - L(\phi_2, \psi_2) - \langle \phi_1 - \phi_2, \nabla_\phi L(\phi_2, \psi_2)\rangle - \langle \psi_1 - \psi_2, \nabla_\psi L(\phi_1, \psi_1)\rangle \geqslant 0$$

$$\forall \quad (\phi_1, \psi_1) \text{ and } (\phi_2, \psi_2) \in E^1 \times F^1 \qquad \text{(IV.2.11)}$$

As $L(\phi, \psi) = M(\phi, \psi) + N(\phi, \psi)$

$\nabla_\phi L(\phi, \psi) = \nabla_\phi M(\phi, \psi) + \nabla_\phi N(\phi, \psi)$

and $\nabla_\psi L(\phi, \psi) = \nabla_\psi M(\phi, \psi) + \nabla_\psi N(\phi, \psi)$, equation (IV.2.11) can be rewritten as

$$\{ M(\phi_1, \psi_1) - M(\phi_2, \psi_2) - \langle \phi_1 - \phi_2, \nabla_\phi M(\phi_2, \psi_2)\rangle - \langle \psi_1 - \psi_2, \nabla_\phi M(\phi_1, \psi_1)\rangle \}$$
$$+ \{ N(\phi_1, \psi_1) - N(\phi_2, \psi_2) - \langle \phi_1 - \phi_2, \nabla_\phi N(\phi_2, \psi_2)\rangle - \langle \psi_1 - \psi_2, \nabla_\psi N(\phi_1, \psi_1)\rangle \} \geqslant 0$$

$$\forall \quad (\phi_1, \psi_1) \text{ and } (\phi_2, \psi_2) \in E^1 \times F^1 \qquad \text{(IV.2.12)}$$

Both terms in curly brackets in equation (IV.2.12) are positive as $M(\phi, \psi)$ and $N(\phi, \psi)$ are both convex/concave saddle functionals; hence equation (IV.2.12) is satisfied and the lemma is proved.

It should be noted that we only require one of $M(\phi, \psi)$ or $N(\phi, \psi)$ to be a strict saddle functional to ensure that $L(\phi, \psi)$ is a strict saddle functional; the criteria for a saddle functional to be strict is that strict inequality holds in equation (IV.2.11) for all $\phi_1 \neq \phi_2$ and $\psi_1 \neq \psi_2$; and provided one of $M(\phi, \psi)$ or $N(\phi, \psi)$ satisfies this criteria, equality in equation (IV.2.12) will only occur for $\phi_1 = \phi_2$ and $\psi_1 = \psi_2$. In fact, it would be enough for one of $M(\phi, \psi)$ and $N(\phi, \psi)$ to be strictly convex in ϕ and the other to be strictly concave in ψ .

Theorem IV.2.1

If $M(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ and $N(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ are convex/concave saddle functionals and the equations $\nabla_\phi L(\phi, \psi) = 0$ and $\nabla_\psi L(\phi, \psi) = 0$ have a solution (ϕ_e, ψ_e), then

$$M(\phi_\beta, \psi_\beta) + N(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq M(\phi_\alpha, \psi_\alpha) + N(\phi_\alpha, \psi_\alpha) \qquad \text{(IV.2.13)}$$

where $\nabla_\phi M(\phi_\beta, \psi_\beta) + \nabla_\phi N(\phi_\beta, \psi_\beta) = 0$ (IV.2.14)

and $\nabla_\psi M(\phi_\alpha, \psi_\alpha) + \nabla_\psi N(\phi_\alpha, \psi_\alpha) = 0$ (IV.2.15)

Proof

Using theorem (II.12.1), as $L(\phi, \psi)$ is a convex/concave saddle functional,

$$L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha) \qquad \text{(IV.2.16)}$$

where $\nabla_\phi L(\phi_\beta, \psi_\beta) = 0$ (IV.2.17)

and $\nabla_\psi L(\phi_\alpha, \psi_\alpha) = 0$ (IV.2.18)

Then substituting equations (IV.2.1), (IV.2.2) and (IV.2.3) into equations (IV.2.16), (IV.2.17) and (IV.2.18) gives the extremum principles specified by equations (IV.2.13), (IV.2.14) and (IV.2.15).

A more general version of the above theorem, the theorem on which the rest of this chapter is based is now given and is followed by its proof and some notes.

Theorem (IV.2.2)

If $M(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ and $N(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ are convex/concave saddle functionals and $L(\phi, \psi) = M(\phi, \psi) + N(\phi, \psi)$ has a stationary value (ϕ_e, ψ_e) given by

$\nabla_\phi L(\phi_e, \psi_e) = 0$ and $\nabla_\psi L(\phi_e, \psi_e) = 0$, then

$$M(\phi_2, \psi_2) + N(\phi_4, \psi_2) - \langle \phi_2 - \phi_4, \nabla_\phi M(\phi_2, \psi_2) \rangle$$
$$\leq L(\phi_e, \psi_e) \leq$$
$$M(\phi_1, \psi_1) + N(\phi_1, \psi_3) - \langle \psi_1 - \psi_3, \nabla_\psi M(\phi_1, \psi_1) \rangle \qquad \text{(IV.2.19)}$$

where ϕ_1, ϕ_2, ϕ_4 and ψ_1, ψ_2, ψ_3 satisfy

$$\nabla_\phi M(\phi_2, \psi_2) + \nabla_\phi N(\phi_4, \psi_2) = 0 \qquad \text{(IV.2.20)}$$

and $\nabla_\psi M(\phi_1, \psi_1) + \nabla_\psi N(\phi_1, \psi_3) = 0$ (IV.2.21)

Proof

To prove the theorem, we need the equations which ensure that $M(\phi, \psi)$ and $N(\phi, \psi)$ are convex/concave saddle functionals; these are

$$M(\phi_a, \psi_a) - M(\phi_b, \psi_b) - \langle \phi_a - \phi_b, \nabla_\phi M(\phi_b, \psi_b) \rangle$$
$$- \langle \psi_a - \psi_b, \nabla_\psi M(\phi_a, \psi_a) \rangle \geq 0$$
$$\forall (\phi_a, \psi_a) \text{ and } (\phi_b, \psi_b) \in E^1 \times F^1 \qquad \text{(IV.2.22)}$$

and

$$N(\phi_c, \psi_c) - N(\phi_d, \psi_d) - \langle \phi_c - \phi_d, \nabla_\phi N(\phi_d, \psi_d) \rangle$$
$$- \langle \psi_c - \psi_d, \nabla_\psi N(\phi_c, \psi_c) \rangle \geq 0$$
$$\forall (\phi_c, \psi_c) \text{ and } (\phi_d, \psi_d) \in E^1 \times F^1 \qquad \text{(IV.2.23)}$$

Lower Bound

Putting $\phi_a = \phi_e$, $\phi_b = \phi_2$, $\phi_c = \phi_e$, $\phi_d = \phi_4$ and $\psi_a = \psi_e$, $\psi_b = \psi_2$, $\psi_c = \psi_e$, $\psi_d = \psi_2$ in equations (IV.2.22) and (IV.2.23), and adding, gives

$$M(\phi_e, \psi_e) - M(\phi_2, \psi_2) - \langle \phi_e - \phi_2, \nabla_\phi M(\phi_2, \psi_2) \rangle$$
$$- \langle \psi_e - \psi_2, \nabla_\psi M(\phi_e, \psi_e) \rangle$$
$$+ N(\phi_e, \psi_e) - N(\phi_4, \psi_2) - \langle \phi_e - \phi_4, \nabla_\phi N(\phi_4, \psi_2) \rangle$$
$$- \langle \psi_e - \psi_2, \nabla_\psi N(\phi_e, \psi_e) \rangle \geq 0$$

or, as $L(\phi_e, \psi_e) = M(\phi_e, \psi_e) + N(\phi_e, \psi_e)$ and

$$\nabla_\phi L(\phi_e, \psi_e) = \nabla_\psi L(\phi_e, \psi_e) = 0$$

$$L(\phi_e, \psi_e) - M(\phi_2, \psi_2) - N(\phi_4, \psi_2)$$
$$- \langle \phi_e - \phi_2, \nabla_\phi M(\phi_2, \psi_2)\rangle - \langle \phi_e - \phi_4, \nabla_\phi N(\phi_4, \psi_2)\rangle \geq 0$$

which can be rewritten

$$L(\phi_e, \psi_e) - M(\phi_2, \psi_2) - N(\phi_4, \psi_2)$$
$$+ \langle \phi_2, \nabla_\phi M(\phi_2, \psi_2)\rangle + \langle \phi_4, \nabla_\phi N(\phi_4, \psi_2)\rangle$$
$$- \langle \phi_e, \nabla_\phi M(\phi_2, \psi_2) + \nabla_\phi N(\phi_4, \psi_2)\rangle \geq 0 \qquad \text{(IV.2.24)}$$

If ϕ_2, ϕ_4 and ψ_2 satisfy $\nabla_\phi M(\phi_2, \psi_2) + \nabla_\phi N(\phi_4, \psi_2) = 0$, then equation (IV.2.24) becomes

$$L(\phi_e, \psi_e) - M(\phi_2, \psi_2) - N(\phi_4, \psi_2) + \langle \phi_2 - \phi_4, \nabla_\phi M(\phi_2, \psi_2)\rangle \geq 0$$

or $L(\phi_e, \psi_e) \geq M(\phi_2, \psi_2) + N(\phi_4, \psi_2)$

$$- \langle \phi_2 - \phi_4, \nabla_\phi M(\phi_2, \psi_2)\rangle$$

as required.

Upper Bound

Similarly, putting $\phi_a = \phi_1$, $\phi_b = \phi_e$, $\phi_c = \phi_1$, $\phi_d = \phi_e$ and

$\psi_a = \psi_1$, $\psi_b = \psi_e$, $\psi_c = \psi_3$, $\psi_d = \psi_e$ in

equations (IV.2.22) and (IV.2.23) and adding, gives

$$M(\phi_1, \psi_1) - M(\phi_e, \psi_e) - \langle \phi_1 - \phi_e, \nabla_\phi M(\phi_e, \psi_e)\rangle$$
$$- \langle \psi_1 - \psi_e, \nabla_\psi M(\phi_1, \psi_1)\rangle$$
$$+ N(\phi_1, \psi_3) - N(\phi_e, \psi_e) - \langle \phi_1 - \phi_e, \nabla_\phi N(\phi_e, \psi_e)\rangle$$
$$- \langle \psi_3 - \psi_e, \nabla_\psi N(\phi_1, \psi_3)\rangle \geq 0$$

or, as $L(\phi_e, \psi_e) = M(\phi_e, \psi_e) + N(\phi_e, \psi_e)$ and

$$\nabla_\phi L(\phi_e, \psi_e) = \nabla_\psi L(\phi_e, \psi_e) = 0.$$

$$L(\phi_e, \psi_e) - M(\phi_1, \psi_1) - N(\phi_1, \psi_3)$$
$$- \langle \phi_e, \nabla_\psi M(\phi_1, \psi_1) + \nabla_\psi N(\phi_1, \psi_3)\rangle$$
$$+ \langle \psi_1, \nabla_\psi M(\phi_1, \psi_1)\rangle + \langle \psi_3, \nabla_\psi N(\phi_1, \psi_3)\rangle \geq 0 \qquad \text{(IV.2.25)}$$

If ϕ_1, ψ_1 and ψ_3 satisfy $\nabla_\psi M(\phi_1, \psi_1) + \nabla_\psi N(\phi_1, \psi_3) = 0$ then equation (IV.2.25) becomes

$L(\phi_e, \psi_e) - M(\phi_1, \psi_1) - N(\phi_1, \psi_3) + \langle \psi_1 - \psi_3, \nabla_\psi M(\phi_1, \psi_1) \rangle \leq 0$

or $L(\phi_e, \psi_e) \leq M(\phi_1, \psi_1) + N(\phi_1, \psi_3) - \langle \psi_1 - \psi_3, \nabla_\psi M(\phi_1, \psi_1) \rangle$

as required.

If $\phi_2 = \phi_4$ and $\psi_1 = \psi_3$, equations (IV.2.19) - (IV.2.21) reduce to the standard dual extremum principles given in theorem (IV.2.1).

If one of $M(\phi, \psi)$ or $N(\phi, \psi)$ is a strict convex/concave saddle functional, (ϕ_e, ψ_e) will be unique.

If $\langle \phi_2 - \phi_4, \nabla_\phi M(\phi_2, \psi_2) \rangle \geqslant 0$, for all ϕ_2, ϕ_4, ψ_2, then the lower bound in equation (IV.2.19) can be reduced to

$$M(\phi_2, \psi_2) + N(\phi_4, \psi_2) \leq L(\phi_e, \psi_e) \qquad \text{(IV.2.26)}$$

and similarly if $\langle \psi_1 - \psi_3, \nabla_\psi M(\phi_1, \psi_1) \rangle \leq 0$ for all ϕ_1, ψ_1 and ψ_3, then the upper bound in equation (IV.2.19) can be reduced to

$$L(\phi_e, \psi_e) \leq M(\phi_1, \psi_1) + N(\phi_1, \psi_3) \qquad \text{(IV.2.27)}$$

Parts of this theory are in (69).

CHAPTER IV

IV.3 Application to the quadratic functional

In this section we are going to apply theorem (IV.2.2) to the usual quadratic functional given by equation (III.1.1).

We start with the strict convex/concave saddle functional defined, as usual, by

$$L(\phi,\psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B\phi\rangle - \tfrac{1}{2}\langle\psi, C\psi\rangle + \langle\phi, f\rangle + \langle\psi, g\rangle \qquad \text{(IV.3.1)}$$

where A is a linear operator with an adjoint A^x such that

$$\langle\phi, A\psi\rangle = \langle\psi, A^x\phi\rangle \qquad \text{(IV.3.2)}$$

and B and C are linear, symmetric, positive-definite operators.

Suppose decompositions of A, B, C, f and g exist such that $A = A_M + A_N$, $B = B_M + B_N$, $C = C_M + C_N$, $f = f_M + f_N$ and $g = g_M + g_N$, where A_M and A_N are linear operators with adjoints $A^x{}_M$ and $A^x{}_N$ satisfying

$$\langle\phi, A_M\psi\rangle = \langle\psi, A^x{}_M\phi\rangle \qquad \text{(IV.3.3)}$$

and $\langle\phi, A_N\psi\rangle = \langle\psi, A^x{}_N\phi\rangle$ (IV.3.4)

and B_M, B_N, C_M and C_N are linear, symmetric, positive-definite operators; an example of a decomposable operator is the operator $-\nabla^2 = -\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2}$,

because if $B = -\nabla^2$ we can take $B_M = -\frac{\partial^2}{\partial x^2}$ and $B_N = -\frac{\partial^2}{\partial y^2}$

If we let

$$M(\phi,\psi) = \langle\phi, A_M\psi\rangle + \tfrac{1}{2}\langle\phi, B_M\phi\rangle - \tfrac{1}{2}\langle\psi, C_M\psi\rangle + \langle\phi, f_M\rangle + \langle\psi, g_M\rangle \qquad \text{(IV.3.5)}$$

and

$$N(\phi,\psi) = \langle\phi, A_N\psi\rangle + \tfrac{1}{2}\langle\phi, B_N\phi\rangle - \tfrac{1}{2}\langle\psi, C_N\psi\rangle + \langle\phi, f_N\rangle + \langle\psi, g_N\rangle \qquad \text{(IV.3.6)}$$

then $L(\phi,\psi) = M(\phi,\psi) + N(\phi,\psi)$ (IV.3.7)

It can easily be shown using theorem (II.9.1) that $M(\phi,\psi)$ and $N(\phi,\psi)$ are convex/concave saddle functionals.

At the stationary point (ϕ_e, ψ_e) of $L(\phi, \psi)$,

$$\nabla_\phi \; L(\phi_e, \psi_e) = A\psi_e + B\phi_e + f = 0 \text{ and} \quad \text{(IV.3.8)}$$

$$\nabla_\psi \; L(\phi_e, \psi_e) = A^x\phi_e - C\psi_e + g = 0 \quad \text{(IV.3.9)}$$

giving $L(\phi_e, \psi_e) = \frac{1}{2}\langle \phi_e, f\rangle + \frac{1}{2}\langle \psi_e, g\rangle$ (IV.3.10)

The gradients of $M(\phi, \psi)$ and $N(\phi, \psi)$ are

$$\nabla_\phi \; M(\phi, \psi) = A_M\psi + B_M\phi + f_M \quad \text{(IV.3.11)}$$

$$\nabla_\psi \; M(\phi, \psi) = A^x{}_M\phi - C_M\psi + g_M \quad \text{(IV.3.12)}$$

$$\nabla_\phi \; N(\phi, \psi) = A_N\psi + B_N\phi + f_N \quad \text{(IV.3.13)}$$

$$\nabla_\psi \; N(\phi, \psi) = A^x{}_N\phi - C_N\psi + g_N \quad \text{(IV.3.14)}$$

Applying theorem (IV.2.2) to the decomposition results in the bounds

$$\begin{aligned}
&\langle \phi_2, A_M\psi_2\rangle + \tfrac{1}{2}\langle \phi_2, B_M\phi_2\rangle - \tfrac{1}{2}\langle \psi_2, C_M\psi_2\rangle \\
&\quad + \langle \phi_2, f_M\rangle + \langle \psi_2, g_M\rangle \\
&+ \langle \phi_4, A_N\psi_2\rangle + \tfrac{1}{2}\langle \phi_4, B_N\phi_4\rangle - \tfrac{1}{2}\langle \psi_2, C_N\psi_2\rangle \\
&\quad + \langle \phi_4, f_N\rangle + \langle \psi_2, g_N\rangle \\
&- \langle \phi_2 - \phi_4, A_M\psi_2 + B_M\phi_2 + f_M\rangle \\
&\qquad\qquad \le L(\phi_e, \psi_e) \le \\
&\langle \phi_1, A_M\psi_1\rangle + \tfrac{1}{2}\langle \phi_1, B_M\phi_1\rangle - \tfrac{1}{2}\langle \psi_1, C_M\psi_1\rangle \\
&\quad + \langle \phi_1, f_M\rangle + \langle \psi_1, g_M\rangle \\
&+ \langle \phi_1, A_N\psi_3\rangle + \tfrac{1}{2}\langle \phi_1, B_N\phi_1\rangle - \tfrac{1}{2}\langle \psi_3, C_N\psi_3\rangle \\
&\quad + \langle \phi_1, f_N\rangle + \langle \psi_3, g_N\rangle \\
&- \langle \psi_1 - \psi_3, A^x{}_M\phi_1 - C_M\psi_1 + g_M\rangle
\end{aligned} \quad \text{(IV.3.15)}$$

where $A_M\psi_2 + B_M\phi_2 + f_M + A_N\psi_2 + B_N\phi_4 + f_N = 0$ (IV.3.16)

and $A^x{}_M\phi_1 - C_M\psi_1 + g_M + A^x{}_N\phi_1 - C_N\psi_3 + g_N = 0$ (IV.3.17)

The last three equations reduce to

$$-\tfrac{1}{2}\langle \phi_2, B_M \phi_2\rangle - \tfrac{1}{2}\langle \phi_4, B_N \phi_4\rangle - \tfrac{1}{2}\langle \psi_2, C\psi_2\rangle + \langle \psi_2, g\rangle$$
$$\le L(\phi_e, \psi_e) \le$$
$$\tfrac{1}{2}\langle \phi_1, B\phi_1\rangle + \tfrac{1}{2}\langle \psi_1, C_M \psi_1\rangle + \tfrac{1}{2}\langle \psi_3, C_N \psi_3\rangle + \langle \phi_1, f\rangle \quad \text{(IV.3.18)}$$

where $A\psi_2 + B_M\phi_2 + B_N\phi_4 + f = 0$ (IV.3.19)

and $A^x\phi_1 - C_M\psi_1 - C_N\psi_3 + g = 0$ (IV.3.20)

If $\phi_2 = \phi_4$ and $\psi_1 = \psi_3$ the above three equations reduce to the classical dual extremum principles given by equations (IV.2.46) - (IV.2.50).

We next show that taking the decomposition operators B_M, B_N, C_M and C_N as multiples of B and C respectively does not lead to improved bounds.

Consider the decompositions

$B_M = \gamma B$, $B_N = (1-\gamma)B$, $0 \le \gamma \le 1$ (IV.3.21)

and $C_M = \delta C$, $C_N = (1-\delta)C$, $0 \le \delta \le 1$ (IV.3.22)

<u>Lower Bound</u>

Let the decomposition lower bound be called L_D; then from equations (IV.3.18), (IV.3.19) and (IV.3.21),

$$L_D = -\frac{\gamma}{2}\langle \phi_2, B\phi_2\rangle - \frac{(1-\gamma)}{2}\langle \phi_4, B\phi_4\rangle$$
$$-\tfrac{1}{2}\langle \psi_2, C\psi_2\rangle + \langle \psi_2, g\rangle \quad \text{(IV.3.23)}$$

where $B(\gamma\phi_2 + (1-\gamma)\phi_4) + A\psi_2 + f = 0$ (IV.3.24)

Using equation (IV.3.24), equation (IV.3.23) can be manipulated to the form

$$L_D = -\tfrac{1}{2}\gamma(1-\gamma)\langle \phi_2 - \phi_4, B(\phi_2 - \phi_4)\rangle$$
$$-\tfrac{1}{2}\langle \psi_2, C\psi_2\rangle + \langle \psi_2, g\rangle$$
$$-\tfrac{1}{2}\langle \gamma\phi_2 + (1-\gamma)\phi_4, B(\gamma\phi_2 + (1-\gamma)\phi_4)\rangle \quad \text{(IV.3.25)}$$

In any given problem the operators B and C and functions f and g are fixed; for given ψ_2, all the terms except the first in equation (IV.3.25) are fixed through equation (IV.3.24), and hence as B is a positive-definite operator and $0 \le \gamma \le 1$, L_D will take its maximum value when $\phi_2 = \phi_4$, giving

$L_D = -\frac{1}{2}\langle\phi_4, B\phi_4\rangle - \frac{1}{2}\langle\psi_2, C\psi_2\rangle + \langle\psi_2, g\rangle$ with $A\psi_2 + B\phi_4 + f = 0$, which are the equations for the classical lower bound. Thus for the decomposition given by equation (IV.3.21), the decomposition lower bound is less than or equal to the classical lower bound.

Upper Bound

Calling the decomposition upper bound $L_D{}^1$ we have, from equations (IV.3.18), (IV.3.20) and (IV.3.22),

$$L_D{}^1 = \frac{\delta}{2}\langle\psi_1, C\psi_1\rangle + \frac{(1-\delta)}{2}\langle\psi_3, C\psi_3\rangle + \tfrac{1}{2}\langle\phi_1, B\phi_1\rangle + \langle\phi_1, f\rangle \qquad \text{(IV.3.26)}$$

where $C(\delta\psi_1 + (1-\delta)\psi_3) - A^x\phi_1 - g = 0$ (IV.3.27)

Using equation (IV.3.27), (IV.3.26) becomes

$$\begin{aligned} L_D{}^1 = {} & \tfrac{1}{2}\delta(1-\delta)\langle\psi_1 - \psi_3, C(\psi_1 - \psi_3)\rangle \\ & + \tfrac{1}{2}\langle\phi_1, B\phi_1\rangle + \langle\phi_1, f\rangle \\ & + \tfrac{1}{2}\langle\delta\psi_1 + (1-\delta)\psi_3, C(\delta\psi_1 + (1-\delta)\psi_3\rangle \end{aligned} \qquad \text{(IV.3.28)}$$

As in the lower bound, $L_D{}^1$ will take its minimum value when $\psi_1 = \psi_3$, for given ϕ_1, giving

$$L_D{}^1 = \tfrac{1}{2}\langle\phi_1, B\phi_1\rangle + \tfrac{1}{2}\langle\psi_3, C\psi_3\rangle + \langle\phi_1, f\rangle$$

where $A^x\phi_1 - C\psi_3 + g = 0$, which are the equations for the classical upper bound. Hence the decomposition given by equation (IV.3.22) does not lead to an upper bound better than the classical.

The rest of this section considers how the decomposition bounds given by equations (IV.3.18), (IV.3.19) and (IV.3.20) can be made to offer an improvement over the classical bounds given in equations (III.2.5) to (III.2.11). Each of the analyses will follow a similar pattern in that conditions on the trial functions which provide improved decomposition bounds will be found by considering the difference between the classical and decomposition bounds.

Lower Bound

From equations (IV.2.14), (IV.2.15), (IV.3.18) and (IV.3.19), noting that $B = B_M + B_N$ and calling the classical bound L_C and the decomposition bound L_D, we have

$$L_C = -\tfrac{1}{2}\langle \phi_s, B_M \phi_s \rangle - \tfrac{1}{2}\langle \phi_s, B_N \phi_s \rangle - \tfrac{1}{2}\langle \psi_s, C\psi_s \rangle + \langle \psi_s, g \rangle \qquad \text{(IV.3.29)}$$

where $A\psi_s + B_M \phi_s + B_N \phi_s + f = 0$ and (IV.3.30)

$$L_D = -\tfrac{1}{2}\langle \phi_2, B_M \phi_2 \rangle - \tfrac{1}{2}\langle \phi_4, B_N \phi_4 \rangle - \tfrac{1}{2}\langle \psi_2, C\psi_2 \rangle + \langle \psi_2, g \rangle \qquad \text{(IV.3.31)}$$

where $A\psi_2 + B_M \phi_2 + B_N \phi_4 + f = 0$ (IV.3.32)

From equations (IV.3.30) and (IV.3.32),

$$A(\psi_2 - \psi_s) + B_M(\phi_2 - \phi_s) + B_N(\phi_4 - \phi_s) = 0 \qquad \text{(IV.3.33)}$$

The decomposition lower bound L_D will better than the classical lower bound if $L_D - L_C > 0$.

$$\begin{aligned}
L_D - L_C = &-\tfrac{1}{2}\langle \phi_2, B_M \phi_2 \rangle + \tfrac{1}{2}\langle \phi_s, B_M \phi_s \rangle \\
&-\tfrac{1}{2}\langle \phi_4, B_N \phi_4 \rangle + \tfrac{1}{2}\langle \phi_s, B_N \phi_s \rangle \\
&-\tfrac{1}{2}\langle \psi_2, C\psi_2 \rangle + \tfrac{1}{2}\langle \psi_s, C\psi_s \rangle \\
&+\langle \psi_2, g \rangle - \langle \psi_s, g \rangle \\
= &\ \tfrac{1}{2}\langle \phi_2 - \phi_s, B_M(\phi_2 - \phi_s) \rangle + \tfrac{1}{2}\langle \phi_4 - \phi_s, B_N(\phi_4 - \phi_s) \rangle \\
&+\tfrac{1}{2}\langle \psi_2 - \psi_s, C(\psi_2 - \psi_s) \rangle \\
&+\langle \psi_2 - \psi_s, A^x \phi_2 + g - C\psi_2 \rangle \\
&+\langle \phi_2 - \phi_s - (\phi_4 - \phi_s), B_N(\phi_4 - \phi_s) \rangle
\end{aligned} \qquad \text{(IV.3.34)}$$

(using equation (IV.3.33)

We want to find conditions on the trial functions ϕ_2, ϕ_4 and ψ_2 which will guarantee that $L_D > L_C$. It cannot be proved from equation (IV.3.34) that $L_D > L_C$ and so we need to put some restrictions on the trial functions.

(a) Assume $\psi_2 = \psi_\beta$ and $\phi_2 = \phi_\beta$; then equations (IV.3.33) and (IV.3.34) become

$$L_D - L_C = \tfrac{1}{2}\langle \phi_4 - \phi_\beta, B_N(\phi_4 - \phi_\beta)\rangle + \langle \phi_2 - \phi_4, B_N(\phi_4 - \phi_\beta)\rangle \qquad (IV.3.35)$$

where $B_N(\phi_4 - \phi_\beta) = 0$ (IV.3.36)

Substituting equation (IV.3.36) into (IV.3.35) gives $L_D - L_C = 0$, and so there is no advantage in assuming $\phi_2 = \phi_\beta$ and $\psi_2 = \psi_\beta$.

Similarly, there is no advantage in assuming $\psi_2 = \psi_\beta$ and $\phi_4 = \phi_\beta$.

(b) Assume $\psi_2 = \psi_\beta$, then equations (IV.3.33) and (IV.3.34) become

$$L_D - L_C = \tfrac{1}{2}\langle \phi_2 - \phi_\beta, B_M(\phi_2 - \phi_\beta)\rangle + \tfrac{1}{2}\langle \phi_4 - \phi_\beta, B_N(\phi_4 - \phi_\beta)\rangle$$
$$+ \langle \phi_2 - \phi_\beta - (\phi_4 - \phi_\beta), B_N(\phi_4 - \phi_\beta)\rangle \qquad (IV.3.37)$$

where $B_M(\phi_2 - \phi_\beta) + B_N(\phi_4 - \phi_\beta) = 0$ (IV.3.38)

$$L_D - L_C = \tfrac{1}{2}\langle \phi_2 - \phi_\beta, B_M(\phi_2 - \phi_\beta)\rangle + \tfrac{1}{2}\langle \phi_4 - \phi_\beta, B_N(\phi_4 - \phi_\beta)\rangle$$
$$+ \langle \phi_2 - \phi_\beta, B_N(\phi_4 - \phi_\beta)\rangle - \langle \phi_4 - \phi_\beta, B_N(\phi_4 - \phi_\beta)\rangle$$
$$= \tfrac{1}{2}\langle \phi_2 - \phi_\beta, B_M(\phi_2 - \phi_\beta)\rangle - \tfrac{1}{2}\langle \phi_4 - \phi_\beta, B_N(\phi_4 - \phi_\beta)\rangle$$
$$- \langle \phi_2 - \phi_\beta, B_M(\phi_2 - \phi_\beta)\rangle$$

(using equation (IV.3.38))

Therefore $L_D - L_C = -\tfrac{1}{2}\langle \phi_2 - \phi_\beta, B_M(\phi_2 - \phi_\beta)\rangle - \tfrac{1}{2}\langle \phi_4 - \phi_\beta, B_N(\phi_4 - \phi_\beta)\rangle$

which is less than zero as B_M and B_N are positive-definite operators; hence there is no advantage to be gained in taking $\phi_2 = \phi_\beta$.

(c) Assume $\phi_2 = \phi_\beta$, then equations (IV.3.33) and (IV.3.34) become

$$L_D - L_C = -\tfrac{1}{2}\langle \phi_4 - \phi_\beta, B_N(\phi_4 - \phi_\beta)\rangle$$
$$+ \tfrac{1}{2}\langle \psi_2 - \psi_\beta, C(\psi_2 - \psi_\beta)\rangle$$
$$+ \langle \psi_2 - \psi_\beta, A^x\phi_2 - C\psi_2 + g\rangle \qquad (IV.3.39)$$

where $A(\psi_2 - \psi_\beta) + B_N(\phi_4 - \phi_\beta) = 0$ (IV.3.40)

Using equation (IV.3.40), (IV.3.39) can be rearranged into

$$L_D - L_C = \tfrac{1}{2}\langle \phi_4 - \phi_\beta , B_N (\phi_4 - \phi_\beta)\rangle + \tfrac{1}{2}\langle \psi_2 - \psi_\beta , C(\psi_2 - \psi_\beta)\rangle$$
$$+\langle \psi_2 - \psi_\beta , A^x\phi_4 - C\psi_2 + g\rangle$$

and so $L_D - L_C$ will be greater than zero if

$$\langle \psi_2 - \psi_\beta , A^x\phi_4 - C\psi_2 + g\rangle \geqslant 0 \qquad \text{(IV.3.41)}$$

where ψ_β is determined from the equation

$$A\psi_\beta + B\phi_2 + f = 0; \qquad \text{(IV.3.42)}$$

thus the condition guaranteeing that the decomposition lower bound is better than the classical lower bound effectively depends only on the decomposition trial functions. Assuming that A^{-1} exists, and can be found easily, then equation (IV.3.41) can be written explicitly in terms of the decomposition trial functions only as

$$\langle \psi_2 + A^{-1}(B\phi_2 + f), A^x\phi_4 - C\psi_2 + g\rangle \geqslant 0 \qquad \text{(IV.3.43)}$$

The flexibility of the decomposition lower bound compared to the classical lower bound can now be seen; once ϕ_β is chosen, then ψ_β is fixed by equation (IV.3.30), but when ϕ_2 is chosen there is still flexibility for the choices of ϕ_4 and ψ_2 by means of equation (IV.3.32) and these can be chosen to satisfy equation (IV.3.41) or (IV.3.43).

(a) Assume $\phi_4 = \phi_\beta$, then equations (IV.3.33) and (IV.3.34) become

$$L_D - L_C = \tfrac{1}{2}\langle \phi_2 - \phi_\beta , B_M (\phi_2 - \phi_\beta)\rangle + \tfrac{1}{2}\langle \psi_2 - \psi_\beta , C(\psi_2 - \psi_\beta)\rangle$$
$$+\langle \psi_2 - \psi_\beta , A^x\phi_2 - C\psi_2 + g\rangle \qquad \text{(IV.3.44)}$$

where $A(\psi_2 - \psi_\beta) + B_M(\phi_2 - \phi_\beta) = 0 \qquad$ (IV.3.45)

$L_D - L_C$ is positive if

$$\langle \psi_2 - \psi_\beta , A^x\phi_2 - C\psi_2 + g\rangle \geqslant 0 \qquad \text{(IV.3.46)}$$

where ψ_β is obtained from the equation

$$A\psi_\beta + B\phi_2 + f = 0 \qquad \text{(IV.3.47)}$$

or, alternatively, assuming A^{-1} exists and can easily be found, equation (IV.3.46) can be written

$$\langle \psi_2 + A^{-1}(B\phi_4 + f), A^x\phi_2 - C\psi_2 + g\rangle \geqslant 0 \qquad \text{(IV.3.48)}$$

An example to show that the conditions can be met will be provided after the upper bound has been considered. Of course, trial functions for L_C and L_D could be chosen without any trial functions in common, and $L_D - L_C$ could still be positive; this would not, however, provide any basis for comparison between the two bounds and therefore will not be considered further.

Upper Bound

From equations (IV.2.49), (IV.2.50), (IV.3.18) and (IV.3.20), noting that $C = C_M + C_N$ and calling the classical bound $L_C{}^1$ and the decomposition bound $L_D{}^1$, we have

$$L_C{}^1 = \tfrac{1}{2}\langle\phi_\alpha, B\phi_\alpha\rangle + \tfrac{1}{2}\langle\psi_\alpha, C_M\psi_\alpha\rangle + \tfrac{1}{2}\langle\psi_\alpha, C_N\psi_\alpha\rangle + \langle\phi_\alpha, f\rangle \qquad \text{(IV.3.49)}$$

where $A^x\phi_\alpha - C_M\psi_\alpha - C_N\psi_\alpha + g = 0$ (IV.3.50)

and $$L_D{}^1 = \tfrac{1}{2}\langle\phi_1, B\phi_1\rangle + \tfrac{1}{2}\langle\psi_1, C_M\psi_1\rangle + \tfrac{1}{2}\langle\psi_3, C_N\psi_3\rangle + \langle\phi_1, f\rangle \qquad \text{(IV.3.51)}$$

where $A^x\phi_1 - C_M\psi_1 - C_N\psi_3 + g = 0$ (IV.3.52)

From equations (IV.3.50) and (IV.3.52),

$$A^x(\phi_1 - \phi_\alpha) - C_M(\psi_1 - \psi_\alpha) - C_N(\psi_3 - \psi_\alpha) = 0 \qquad \text{(IV.3.53)}$$

The decomposition upper bound $L_D{}^1$ will be better than the classical upper bound $L_C{}^1$ if $L_C{}^1 - L_D{}^1 > 0$.

Using equation (IV.3.52),

$$\begin{aligned} L_C{}^1 - L_D{}^1 = {} & \tfrac{1}{2}\langle\phi_1 - \phi_\alpha, B(\phi_1 - \phi_\alpha)\rangle + \tfrac{1}{2}\langle\psi_1 - \psi_\alpha, C_M(\psi_1 - \psi_\alpha)\rangle \\ & + \tfrac{1}{2}\langle\psi_3 - \psi_\alpha, C_N(\psi_3 - \psi_\alpha)\rangle \\ & - \langle\phi_1 - \phi_\alpha, A\psi_1 + B\phi_1 + f\rangle \\ & + \langle\psi_1 - \psi_\alpha - (\psi_3 - \psi_\alpha), C_N(\psi_3 - \psi_\alpha)\rangle \end{aligned} \qquad \text{(IV.3.54)}$$

As in the lower bound, taking (a) $\phi_1 = \phi_\alpha$ and $\psi_1 = \psi_\alpha$ or (b) $\phi_1 = \phi_\alpha$ and $\psi_3 = \psi_\alpha$ or (c) $\phi_1 = \phi_\alpha$ does not offer any improvement over the classical upper bound.

(d) Assume $\psi_1 = \psi_\alpha$, then equations (IV.3.54) and (IV.3.53) become

$$L_C^1 - L_D^1 = \tfrac{1}{2}\langle \phi_1 - \phi_\alpha , B(\phi_1 - \phi_\alpha)\rangle - \tfrac{1}{2}\langle \psi_3 - \psi_\alpha , C_N (\psi_3 - \psi_\alpha)\rangle - \langle \phi_1 - \phi_\alpha , A\psi_1 + B\phi_1 + f\rangle \quad (IV.3.55)$$

where $A^x(\phi_1 - \phi_\alpha) - C_N(\psi_3 - \psi_\alpha) = 0$ (IV.3.56)

Using equation (IV.3.56), (IV.3.55) becomes

$$L_C^1 - L_D^1 = \tfrac{1}{2}\langle \phi_1 - \phi_\alpha , B(\phi_1 - \phi_\alpha)\rangle + \tfrac{1}{2}\langle \psi_3 - \psi_\alpha , C_N (\psi_3 - \psi_\alpha)\rangle - \langle \phi_1 - \phi_\alpha , A\psi_3 + B\phi_1 + f\rangle \quad (IV.3.57)$$

$L_C^1 - L_D^1$ will be positive if

$$\langle \phi_1 - \phi_\alpha , A\psi_3 + B\phi_1 + f\rangle \leq 0 \quad (IV.3.58)$$

where ϕ_α is obtained from the equation

$$A^x \phi_\alpha - C\psi_1 + g = 0 \quad (IV.3.59)$$

or, alternatively, assuming that $(A^x)^{-1}$ exists and can be easily found, equation (IV.3.58) can be written

$$\langle \phi_1 - (A^x)^{-1}(C\psi_1 - g), A\psi_3 + B\phi_1 + f\rangle \leq 0 \quad (IV.3.60)$$

(e) Finally, assume that $\psi_3 = \psi_\alpha$, then equations (IV.3.54) and (IV.3.53) become

$$L_C^1 - L_D^1 = \tfrac{1}{2}\langle \phi_1 - \phi_\alpha , B(\phi_1 - \phi_\alpha)\rangle + \tfrac{1}{2}\langle \psi_1 - \psi_\alpha , C_M (\psi_1 - \psi_\alpha)\rangle - \langle \phi_1 - \phi_\alpha , A\psi_1 + B\phi_1 + f\rangle \quad (IV.3.61)$$

where $A^x(\phi_1 - \phi_\alpha) - C_M(\psi_1 - \psi_\alpha) = 0$ (IV.3.62)

$L_C^1 - L_D^1$ is positive if $\langle \phi_1 - \phi_\alpha , A\psi_1 + B\phi_1 + f\rangle \leq 0$ (IV.3.63)

where ϕ_α is determined by $A^x\phi_\alpha - C\psi_3 + g = 0$ (IV.3.64)

or, assuming that $(A^{-1})^x$ exists and can easily be found, $L_C^1 - L_D^1 > 0$ if

$$\langle \phi_1 - (A^x)^{-1}(C\psi_3 - g), A\psi_1 + B\phi_1 + f\rangle \leq 0 \quad (IV.3.65)$$

We end this section with an example to illustrate that it is possible to satisfy the conditions which guarantee that the decomposition bounds are better than the classical bounds.

The method will be applied to the problem looked at earlier in the context of iterative methods in section (III.9):

$$(K + I)\phi(x) = x^2 - 2x, \quad x \in [0,1]$$

where $K\phi(x) = \int_0^x k(x,y)\phi(y)\,dy + \int_x^1 k(x,y)\phi(y)\,dy$ (IV.3.66)

and $k(x,y) = \begin{cases} y, & x \geq y \\ x, & x \leq y \end{cases} \quad x, y \in [0,1]$

and I is the usual identity operator.

The inner product is given by the equation

$$\langle u,v \rangle = \int_0^1 u(x)\, v(x)\, dx \tag{IV.3.67}$$

In equation (IV.3.1) let us specify A, A^x, B, C, f and g as:

$$B = I + 2K, \quad C = A = A^x = I, \quad f = 2(2x - x^2) \text{ and } g = 0 \tag{IV.3.68}$$

Then equations (IV.3.8) and (IV.3.9) become

$\psi_e + (I + 2K)\psi_e + 2(2x - x^2) = 0$ and $\phi_e - \psi_e = 0$ or

$2(I + K)\phi_e = 2(x^2 - 2x)$ or $(I + K)\phi_e = x^2 - 2x$ as required.

Now let the operators in equations (IV.3.5) and (IV.3.6) be given by

$$A_M = A = I \text{ and } A_N = 0; \tag{IV.3.69}$$

$$B_M = B - m^2 I = (I + 2K - m^2 I) \text{ and } B_N = m^2 I, \tag{IV.3.70}$$

where m^2 is a real non-zero number and $B - m^2 I$ is positive;

$$C_M = C - n^2 I = (1 - n^2) I \text{ and } C_N = n^2 I \tag{IV.3.71}$$

where n^2 is a real non-zero number and $C - n^2 I$ is positive;

$$f_M = f = 2(2x - x^2) \text{ and } f_N = 0; \tag{IV.3.72}$$

$$\text{and } g_M = g_N = 0 \tag{IV.3.73}$$

Lower Bound

From equations (III.2.6), (III.2.7) and (IV.3.68), we have the classical lower bound

$$L_C = -\tfrac{1}{2}\int_0^1 \{\phi_\beta^{\,2} + 2\,\phi_\beta\, K\phi_\beta + \psi_\beta^{\,2}\}\, dx \tag{IV.3.74}$$

where $\psi_{\beta} + \phi_{\beta} + 2K\phi_{\beta} + 2(2x - x^2) = 0$ (IV.3.75)

and, from equations (IV.3.18), (IV.3.19), (IV.3.68) and (IV.3.70), the decomposition lower bound is

$$L_D = -\tfrac{1}{2}\int_0^1 \left\{(1 - m^2)\phi_2^2 + 2\phi_2 K\phi_2 + m^2\phi_4^2 + \psi_2^2\right\} dx \qquad \text{(IV.3.76)}$$

where $\psi_2 + (1 - m^2)\phi_2 + 2K\phi_2 + m^2\phi_4 + 2(2x - x^2) = 0$ (IV.3.77)

If we choose $\phi_2 = \phi_{\beta}$, L_D will be greater than L_C if ϕ_2, ϕ_4 and ψ_2 are chosen so that

$$\int_0^1 (\phi_2 + \psi_2 + 2K\phi_2 + 2(2x - x^2)) (\phi_4 - \psi_2)) \, dx \geqslant 0$$

(using equations (IV.3.43) and (IV.3.68)); or, using equation (IV.3.72), ϕ_2, ϕ_4 and ψ_2 should be chosen so that

$$m^2 \int_0^1 (\phi_2 - \phi_4)(\phi_4 - \psi_2) \, dx \geqslant 0 \qquad \text{(IV.3.78)}$$

<u>Sample trial functions</u>

Let $\phi_{\beta} = 1$; then $\psi_{\beta} = 2(x^2 - 2x) - 1 - 2K(1)$ or, using the formula for $K(x^n)$ from appendix IV,

$\psi_{\beta} = 2(x^2 - 2x) - 1 - 2(x - \frac{x^2}{2}) = 3x^2 - 6x - 1$. This gives

$$L_C = -\tfrac{1}{2}\int_0^1 \left\{1 + 2(x - \frac{x^2}{2}) + (3x^2 - 6x - 1)^2\right\} dx$$

$$= \frac{-86}{15}$$

Let $\phi_2 = 1$; then we want ϕ_4 and ψ_2 such that

$$(1 - m^2) + 2(x - \frac{x^2}{2}) + \psi_2 + m^2\phi_4 + 2(2x - x^2) = 0 \qquad \text{(IV.3.79)}$$

and $$m^2 \int_0^1 (1 - \phi_4)(\phi_4 - \psi_2) \, dx \geqslant 0 \qquad \text{(IV.3.80)}$$

Equations (IV.3.79) and (IV.3.80) are satisfied if we take

$\phi_4 = 0$ and $\psi_2 = 3x^2 - 6x - 1 + m^2$, giving

$$m^2 \int_0^1 (1 - \phi_4)\,(\phi_4 - \psi_2)\,dx = m^2 \int_0^1 -\,(3x^2 - 6x - 1 + m^2)\,dx$$

$$= m^2\,(3 - m^2). \qquad \text{(IV.3.81)}$$

We require m^2 such that $B - m^2 I \geqslant 0$ or $(I + 2K - m^2 I)$ is a positive operator. As $K > 0$ (from example 1 of section II.15), $(I + 2K - m^2 I) \geqslant 0$ if $1 - m^2 \geqslant 0$ or $m^2 \leqslant 1$, which implies that $3 - m^2$ is positive and hence $m^2 \int_0^1 (1 - \phi_4)\,(\phi_4 - \psi_2)\,dx$ is positive.

With ϕ_4 and ψ_2 as above,

$$L_D = -\tfrac{1}{2} \int_0^1 \left\{ (1 - m^2) + 2\left(x - \frac{x^2}{2}\right) + (3x^2 - 6x - 1 + m^2) \right\} dx$$

$$= -\frac{86}{15} + \frac{7m^2}{2} - \frac{m^4}{2} \qquad \text{(IV.3.82)}$$

This gives $L_D - L_C = \dfrac{m^2}{2}\,(7 - m^2)$, which is greater than zero and hence L_D is a better bound than L_C. L_D can be written as

$L_D = -\tfrac{1}{2}\left(m^2 - \dfrac{7}{2}\right)^2 + \dfrac{47}{120}$, which has its maximum when $m^2 = \dfrac{7}{2}$; this is not allowable as we require $m^2 \leqslant 1$. Therefore taking the highest possible value for m^2, that is $m^2 = 1$, then we have $L_D = -\dfrac{41}{15}$.

Upper Bound

From equations (III.2.8), (III.2.9) and (IV.3.68) the classical upper bound is

$$L_C{}^1 = \int_0^1 \left\{ \tfrac{1}{2}\,\phi_\alpha{}^2 + \phi_\alpha K \phi_\alpha + \tfrac{1}{2}\,\psi_\alpha{}^2 + 2(2x - x^2)\,\phi_\alpha \right\} dx \qquad \text{(IV.3.83)}$$

where $\phi_\alpha - \psi_\alpha = 0$ (IV.3.84)

and, from equations (IV.3.18), (IV.3.20), (IV.3.68) and (IV.3.71), the decomposition upper bound is

$$L_D{}^1 = \int_0^1 \left\{ \tfrac{1}{2}\phi_1{}^2 + \phi_1 K \phi_1 + \tfrac{1}{2}\,(1 - n^2)\,\psi_1{}^2 + \tfrac{1}{2}\,n^2 \psi_3{}^2 \right.$$

$$\left. + 2\,\phi_1\,(2x - x^2) \right\} dx \qquad \text{(IV.3.85)}$$

where $\phi_1 - (1 - n^2)\,\psi_1 - n^2 \psi_3 = 0$ (IV.3.86)

If we choose $\phi_1 = \psi_\alpha$, L_D^1 will be less than L_C^1 if ϕ_1, ψ_1 and ψ_3 are chosen so that

$$\int_0^1 \left\{ (\phi_1 - \psi_1)\ (\psi_3 + \phi_1 + 2K\phi_1 + 2(2x - x^2) \right\} dx \leq 0$$

(using equations (IV.3.60) and (IV.3.68); or, using equation (IV.3.86), ϕ_1, ψ_1 and ψ_3 should be chosen so that

$$n^2 \int_0^1 \left\{ (\psi_1 - \psi_3)\ (\psi_3 + \phi_1 + 2K\phi_1 + 2(2x - x^2)) \right\} dx \geq 0 \qquad \text{(IV.3.87)}$$

<u>Sample trial functions</u>

Let $\psi_\alpha = 1$; then $\phi_\alpha = 1$ and so

$$L_C^1 = \int_0^1 \left\{ \tfrac{1}{2} + (x - \frac{x^2}{2}) + \tfrac{1}{2} + 2(2x - x^2 \right\} dx$$

$$= \frac{8}{3}$$

Let $\psi_1 = 1$; then we want to find ϕ_1 and ψ_3 such that

$$\phi_1 - (1 - n^2) - n^2\psi_3 = 0 \qquad \text{(IV.3.88)}$$

$$\text{and } n^2 \int_0^1 \left\{ (1 - \psi_3)\ (\psi_3 + \phi_1 + 2K\phi_1 + 2(2x - x^2)) \right\} dx \geq 0 \qquad \text{(IV.3.89)}$$

These two equations are satisfied if we take

$\psi_3 = 0$ and $\phi_1 = 1 - n^2$, for then equation (IV.3.89) becomes

$$n^2 \int_0^1 \left\{ 1 - n^2 + 2(1 - n^2)\ (x - \frac{x^2}{2}) + 2(2x - x^2) \right\} dx \geq 0$$

or $n^2 (3 - \frac{5}{3} n^2) \geq 0$, which is true, as, from equation (IV.3.71), to make $C - n^2 I$ positive we require $n^2 \leq 1$.

We then have

$$L_D^1 = \tfrac{1}{2}\int_0^1 \left\{ (1 - n^2) + 2(1 - n^2)\ (x - \frac{x^2}{2}) + 1 \right.$$

$$\left. + 4(1 - n^2)\ (2x - x^2) - n^2 \right\} dx$$

$$= \frac{8}{3} + \frac{1}{3} n^4 - 3n^2$$

and thus $L_C^1 - L_D^1 = 3n^2 \left(1 - \frac{1}{9} n^2\right)$ which is greater than zero so that L_D^1 is a better bound than L_C^1.

Writing $L_D^1 = \frac{1}{3}\left(n^2 - \frac{9}{2}\right)^2 - \frac{49}{12}$, it can be seen that the minimum of L_D^1 occurs when $n^2 = \frac{9}{2}$, which is not permissible. We therefore take the highest possible value for n^2, that is $n^2 = 1$, which gives $L_D^1 = 0$.

The results are given in table (IV.3.1):

Table IV.3.1

	Classical D.E.P	Decomposition D.E.P
Lower Bound	$-\frac{86}{15}$	$-\frac{41}{15}$
Upper Bound	$\frac{8}{3}$	0

Using the decomposition dual extremum principles has resulted in sharper bounds than those obtained using classical dual extremum principles. Obviously, as the trial functions used were very simple, the bounds are not very close together; they could be made sharper by the application of iterative methods. The combination of decomposition dual extremum principles and iterative methods will be considered in section IV.5.

IV.4 Convergence for iterative schemes using classical dual extremum principles

In this section we consider conditions on the operators in the usual quadratic functional which ensure that the cobweb iterative schemes given in section III.3 converge. The conditions for schemes A and B are explicitly developed; it is not necessary to do the same for schemes C and D, as cobweb iterative scheme A is the same as cobweb iterative scheme C, and similarly schemes B and D are the same. At the end of the section the results are listed, together with the conditions for convergence for the four optimising iterative schemes given in section III.9; these are then used for comparison purposes in the next section.

We start with the functional

$$L(\phi,\psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B\phi\rangle - \tfrac{1}{2}\langle\psi, C\psi\rangle + \langle\phi, f\rangle + \langle\psi, g\rangle \qquad \text{(IV.4.1)}$$

where the linear operator A has a linear adjoint A^x such that $\langle\phi, A\psi\rangle = \langle\psi, A^x\phi\rangle$, and the linear operators B and C are symmetric and positive definite. The functional derivatives of $L(\phi,\psi)$ are given by

$$\nabla_\phi L(\phi,\psi) = A\psi + B\phi + f \qquad \text{(IV.4.2)}$$

and $$\nabla_\psi L(\phi,\psi) = A^x\phi - C\psi + g \qquad \text{(IV.4.3)}$$

Setting $\nabla_\phi L(\phi,\psi) = \nabla_\psi L(\phi,\psi) = 0$ gives the unique saddle point (ϕ_e, ψ_e) which is the solution of

$$A\psi_e + B\phi_e + f = 0 \qquad \text{(IV.4.4)}$$

and $$A^x\phi_e - C\psi_e + g = 0 \qquad \text{(IV.4.5)}$$

By Lemma (II.18.4), B^{-1} and C^{-1} exist if B and C are bounded below. We are going to assume that A and A^x have inverses A^{-1} and $(A^x)^{-1}$ respectively. The convergence of each iterative scheme is shown by the use of theorem (II.16.1), which, for clarity, is reproduced here.

CHAPTER IV

Theorem (II.16.1)

If P is a linear, self-adjoint operator and there exist real numbers q and Q such that

$$-I < qI \leq P \leq QI < I \qquad \text{(IV.4.6)}$$

then the iterative scheme defined by the equation

$$u_{n+1} = Pu_n, \; n = 0,1,2,\ldots \qquad \text{(IV.4.7)}$$

converges; that is, $\lim_{n \to \infty} \|u_{n+1}\| = 0$ (IV.4.8)

P can be a positive or a negative operator.

What we need to show, for each cobweb iteration, is that

$$\lim_{n \to \infty} \|\phi_{n+1} - \phi_e\| + \|\psi_{n+1} - \psi_e\| = 0 \qquad \text{(IV.4.9)}$$

If $\psi_{n+1} - \psi_e$ can be expressed in terms of $\phi_{n+1} - \phi_e$, that is $\psi_{n+1} - \psi_e = R(\phi_{n+1} - \phi_e)$ for some operator R, then equation (IV.4.9) is satisfied if we show that

$$\lim_{n \to \infty} \|\phi_{n+1} - \phi_e\| = 0 \qquad \text{(IV.4.10)}$$

The technique is to rearrange the equations defining the iteration, together with equations (IV.4.4) and (IV.4.8), into the form

$$\phi_{n+1} - \phi_e = P(\phi_n - \phi_e) \qquad \text{(IV.4.11)}$$

Then, taking $u_n = \phi_n - \phi_e$, we can conclude that equation (IV.4.10) (and thus equation (IV.4.9)) is satisfied if P satisfies equation (IV.4.6).

It also follows from section III.3 that the sequences of upper and lower bounds obtained from the iterations converge to the stationary value $L(\phi_e, \psi_e)$.

Iteration A

From equations (III.3.1), (IV.4.2) and (IV.4.3) this is defined as:

$$A\psi_n + B\phi_n + f = 0 \qquad \text{(IV.4.12)}$$

$$A^x\phi_{n+1} - C\psi_n + g = 0 \qquad \text{(IV.4.13)}$$

$$A\psi_{n+1} + B\phi_{n+1} + f = 0 \qquad \text{(IV.4.14)}$$

Theorem (II.16.1)

If P is a linear, self-adjoint operator and there exist real numbers q and Q such that

$$- I < q I \leq P \leq QI < I \tag{IV.4.6}$$

then the iterative scheme defined by the equation

$$u_{n+1} = Pu_n, \; n = 0,1,2,\ldots \tag{IV.4.7}$$

converges; that is, $\lim_{n \to \infty} \| u_{n+1} \| = 0$ (IV.4.8)

P can be a positive or a negative operator.

What we need to show, for each cobweb iteration, is that

$$\lim_{n \to \infty} \| \phi_{n+1} - \phi_e \| + \| \psi_{n+1} - \psi_e \| = 0 \tag{IV.4.9}$$

If $\psi_{n+1} - \psi_e$ can be expressed in terms of $\phi_{n+1} - \phi_e$, that is $\psi_{n+1} - \psi_e = R(\phi_{n+1} - \phi_e)$ for some operator R, then equation (IV.4.9) is satisfied if we show that

$$\lim_{n \to \infty} \| \phi_{n+1} - \phi_e \| = 0 \tag{IV.4.10}$$

The technique is to rearrange the equations defining the iteration, together with equations (IV.4.4) and (IV.4.8), into the form

$$\phi_{n+1} - \phi_e = P(\phi_n - \phi_e) \tag{IV.4.11}$$

Then, taking $u_n = \phi_n - \phi_e$, we can conclude that equation (IV.4.10) (and thus equation (IV.4.9)) is satisfied if P satisfies equation (IV.4.6).

It also follows from section III.3 that the sequences of upper and lower bounds obtained from the iterations converge to the stationary value $L(\phi_e, \psi_e)$.

Iteration A

From equations (III.3.1), (IV.4.2) and (IV.4.3) this is defined as:

$$A\psi_n + B\phi_n + f = 0 \tag{IV.4.12}$$

$$A^x\phi_{n+1} - C\psi_n + g = 0 \tag{IV.4.13}$$

$$A\psi_{n+1} + B\phi_{n+1} + f = 0 \tag{IV.4.14}$$

Using the above three equations and equations (IV.4.4) and (IV.4.5), we have

$$A^x (\phi_{n+1} - \phi_e) = C(\psi_n - \psi_e) \quad (IV.4.15)$$

$$A (\psi_n - \psi_e) = - B(\phi_n - \phi_e) \quad (IV.4.16)$$

$$A (\psi_{n+1} - \psi_e) = - B(\phi_{n+1} - \phi_e) \quad (IV.4.17)$$

Hence

$$A^x (\phi_{n+1} - \phi_e) = CA^{-1} A (\psi_n - \psi_e)$$

$$= - CA^{-1} B (\phi_n - \phi_e) \text{ from (IV.4.16)}$$

gives $\phi_{n+1} - \phi_e = -(A^x)^{-1} CA^{-1} B (\phi_n - \phi_e)$ (IV.4.18)

Also $\psi_{n+1} - \psi_e = -A^{-1} B (\phi_{n+1} - \phi_e)$ (IV.4.19)

Therefore the iteration converges provided $P = -(A^x)^{-1} CA^{-1} B$ satisfies equation (IV.4.6).

<u>Iteration B</u>

From equations (III.3.2), (IV.4.4) and (IV.4.5), we have

$$A\psi_n + B\phi_n + f = 0 \quad (IV.4.20)$$

$$A^x\phi_n - C\psi_{n+1} + g = 0 \quad (IV.4.21)$$

$$A\psi_{n+1} + B\phi_{n+1} + f = 0 \quad (IV.4.22)$$

Using these three equations and equations (IV.4.4) and (IV.4.5), we have

$$B(\phi_{n+1} - \phi_e) = - A(\psi_{n+1} - \psi_e) \quad (IV.4.23)$$

$$C(\psi_{n+1} - \psi_e) = A^x (\phi_n - \phi_e) \quad (IV.4.24)$$

$$B(\phi_n - \phi_e) = - A (\psi_n - \psi_e) \quad (IV.4.25)$$

Hence

$$B(\phi_{n+1} - \phi_e) = - AC^{-1} C (\psi_{n+1} - \psi_e)$$

$$= - AC^{-1} A^x (\phi_n - \phi_e) \text{ from (IV.4.24)}$$

giving $\phi_{n+1} - \phi_e = - B^{-1} AC^{-1} A^x (\phi_n - \phi_e)$ (IV.4.26)

Also, $\psi_{n+1} - \psi_e = - A^{-1} B (\phi_{n+1} - \phi_e)$ (IV.4.27)

Therefore the iteration converges provided $P = - B^{-1} AC^{-1} A^x$ satisfies equation (IV.4.6).

These results, and the conditions for convergence for the optimising iterative schemes from section III.9, are tabulated in Tables (IV.4.1) and (IV.4.2).

Table (IV.4.1) : Cobweb Iterative Schemes

In each iteration, $\lim_{n \to \infty} \quad \|\phi_{n+1} - \phi_e\| \quad + \quad \|\psi_{n+1} - \psi_e\| \quad = 0$ if the given operator P is linear, and self-adjoint, and satisfies $-I < qI \leq P \leq QI < I$ for some real numbers q and Q.

Iteration A	Iteration B	Iteration C	Iteration D
$A\psi_n + B\phi_n + f = 0$	$A\psi_n + B\phi_n + f = 0$	$A^x\phi_n - C\psi_n + g = 0$	$A^x\phi_n - C\psi_n + g = 0$
$A^x\phi_{n+1} - C\psi_n + g = 0$	$A^x\phi_n - C\psi_{n+1} + g = 0$	$A\psi_{n+1} + B\phi_n + f = 0$	$A\psi_n + B\phi_{n+1} + f = 0$
$A\psi_{n+1} + B\phi_{n+1} + f = 0$	$A\psi_{n+1} + B\phi_{n+1} + f = 0$	$A^x\phi_{n+1} - C\psi_{n+1} + g = 0$	$A^x\phi_{n+1} - C\psi_{n+1} + g = 0$
$P = -(A^x)^{-1} CA^{-1} B$	$P = -B^{-1} AC^{-1} A^x$	$P = -(A^x)^{-1} CA^{-1} B$	$P = -B^{-1} AC^{-1} A^x$

Table (IV.4.2) : Optimising Iterative Schemes

In each iteration, the defining equations are those given together with $\phi_{2n+2} = \lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$ and $\psi_{2n+2} = \lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$ where λ_{2n+2} is also specified

$\lim_{n \to \infty} \|\phi_{2n+2} - \phi_e\| + \|\psi_{2n+2} - \psi_e\| = 0$ if the operators A, A^x, B and C satisfy the given conditions.

	Iteration A	Iteration B	Iteration C	Iteration D
Specification	$A\psi_{2n} + B\phi_{2n} + f = 0$	$A\psi_{2n} + B\phi_{2n} + f = 0$	$A^x\phi_{2n} - C\psi_{2n} + g = 0$	$A^x\phi_{2n} - C\psi_{2n} + g = 0$
	$A^x\phi_{2n+1} - C\psi_{2n} + g = 0$	$A^x\phi_{2n} - C\psi_{2n+1} + g = 0$	$A\psi_{2n+1} + B\phi_{2n} + f = 0$	$A\psi_{2n} + B\phi_{2n+1} + f = 0$
	$A\psi_{2n+1} + B\phi_{2n+1} + f = 0$	$A\psi_{2n+1} + B\phi_{2n+1} + f = 0$	$A^x\phi_{2n+1} - C\psi_{2n+1} + g = 0$	$A^x\phi_{2n+1} - C\psi_{2n+1} + g = 0$
λ_{2n+2}	$\dfrac{\langle\psi_{2n}-\psi_{2n+1}, C(\psi_{2n}-\psi_{2n+1})\rangle}{\{\langle\phi_{2n}-\phi_{2n+1}, B(\phi_{2n}-\phi_{2n+1})\rangle + \langle\psi_{2n}-\psi_{2n+1}, C(\psi_{2n}-\psi_{2n+1})\rangle\}}$	$\dfrac{\langle\phi_{2n}-\phi_{2n+1}, B(\phi_{2n}-\phi_{2n+1})\rangle}{\{\langle\phi_{2n}-\phi_{2n+1}, B(\phi_{2n}-\phi_{2n+1})\rangle + \langle\psi_{2n}-\psi_{2n+1}, C(\psi_{2n}-\psi_{2n+1})\rangle\}}$	$\dfrac{\langle\phi_{2n}-\phi_{2n+1}, B(\phi_{2n}-\phi_{2n+1})\rangle}{\{\langle\phi_{2n}-\phi_{2n+1}, B(\phi_{2n}-\phi_{2n+1})\rangle + \langle\psi_{2n}-\psi_{2n+1}, C(\psi_{2n}-\psi_{2n+1})\rangle\}}$	$\dfrac{\langle\psi_{2n}-\psi_{2n+1}, C(\psi_{2n}-\psi_{2n+1})\rangle}{\{\langle\phi_{2n}-\phi_{2n+1}, B(\phi_{2n}-\phi_{2n+1})\rangle + \langle\psi_{2n}-\psi_{2n+1}, C(\psi_{2n}-\psi_{2n+1})\rangle\}}$
Conditions for Convergence	$A^x B^{-1} A = I$ and C is bounded	$C = I$; $A^x B^{-1} A$ is linear and self-adjoint and $(A^x B^{-1} A + I)$ is positive-definite	$A C^{-1} A^x = I$ and B is bounded	$B = I$; $AC^{-1} A^x$ is linear and self-adjoint and $(AC^{-1} A^x + I)$ is positive-definite

As can be seen from the two tables, the conditions for convergence in the cobweb iterative scheme are generally more restrictive than the convergence conditions in the optimising iterative schemes; for instance, if we choose $A = A^x = B = I$, then for cobweb iterative scheme A to converge we require C to satisfy the conditions $-I < qI \leq C \leq QI < I$, but for the optimising iterative scheme A to converge we just require that C is a bounded operator.

IV.5 Convergence for iterative schemes using decomposition dual extremum principles

In this section we are going to apply cobweb and optimising iterative methods to the equations obtained when the decomposition given in section IV.3 is used. Conditions for convergence for general operators A, A^x, B and C will be found, using the convergence conditions for bounded operators; as we wish to make comparisons between schemes, convergence conditions for unbounded operators will not be considered.

From equations (IV.3.18) to (IV.3.20), the decomposition dual extremum principles are

$$- \tfrac{1}{2}\langle \phi_2, B_M\phi_2\rangle - \tfrac{1}{2}\langle \phi_4, B_N\phi_4\rangle - \tfrac{1}{2}\langle \psi_2, C\psi_2\rangle + \langle \psi_2, g\rangle$$
$$\leq L(\phi_e, \psi_e) \leq$$
$$\tfrac{1}{2}\langle \phi_1, B\phi_1\rangle + \tfrac{1}{2}\langle \psi_1, C_M\psi_1\rangle + \tfrac{1}{2}\langle \psi_3, C_N\psi_3\rangle + \langle \phi_1, f\rangle \qquad (IV.5.1)$$

where $A\psi_2 + B_M\phi_2 + B_N\phi_4 + f = 0$ (IV.5.2)

and $A^x\phi_1 - C_M\psi_1 - C_N\psi_3 + g = 0$; (IV.5.3)

(ϕ_e, ψ_e) is the unique solution of

$$A\psi_e + B\phi_e + f = 0 \qquad (IV.5.4)$$

and $A^x\phi_e - C\psi_e + g = 0$ (IV.5.5)

or, as $B = B_M + B_N$ and $C = C_M + C_N$,

$$A\psi_e + B_M\phi_e + B_N\phi_e + f = 0 \qquad (IV.5.6)$$

and $A^x\phi_e - C_M\psi_e - C_N\psi_e + g = 0$ (IV.5.7)

As usual, we assume that A^{-1} and $(A^x)^{-1}$ exist, as well as B_M^{-1}, B_N^{-1}, C_M^{-1}, C_N^{-1}, B^{-1} and C^{-1}.

In order to make comparisons between the iterative schemes given for the classical dual extremum principles in section (IV.4) and the iterative schemes developed here for the decomposition dual extremum principles, the trial functions ϕ_1, ϕ_2, ϕ_4 and ψ_1, ψ_2, ψ_3 will be chosen so that one of equations (IV.5.2) and (IV.5.3) just contains ϕ_n and ψ_n (or ϕ_{2n} and ψ_{2n}

for the optimising iterative schemes) and the other equation contains ϕ_n, ψ_n and ϕ_{n+1} or ϕ_n, ψ_n and ψ_{n+1} (or ϕ_{2n}, ψ_{2n} and ϕ_{2n+1} or ϕ_{2n}, ψ_{2n} and ψ_{2n+1}, for the optimising iterative schemes).

Because of the inclusion of the extra trial functions we cannot directly use the cobweb iterative schemes given in equations (III.3.1) to (III.3.4), or the optimising iterative schemes given in theorems (III.7.1), (III.7.2), (III.8.1) and (III.8.2); instead we exploit the inclusion of the extra trial functions by choosing either ϕ_2 different from ϕ_4 with ϕ_1 equal to ϕ_2 or ϕ_4, or ψ_1 different from ψ_3 with ψ_2 equal to ψ_1 or ψ_3.

A consequence of the conditions in the last two paragraphs is that we cannot find iterative schemes analogous to cobweb and optimising iterative schemes A and C; for instance, if we chose $\psi_2 = \psi_{n+1}$ we would also have to choose $\phi_2 = \phi_4 = \phi_1 = \phi_n$ and $\psi_1 = \psi_3 = \psi_n$, which just results in the classical cobweb iterative scheme.

Other choices of trial functions leading to different cobweb iterative schemes can be specified, but as we are only interested in schemes which can be compared with those obtained using classical dual extremum principles the other possible choices of trial functions will not be pursued.

As we can choose B_M and B_N arbitrarily, provided $B_M + B_N = B$, and similarly we can choose C_M and C_N arbitrarily, where $C_M + C_N = C$, there is no essential difference between choosing $\phi_2 = \phi_n$ and $\phi_4 = \phi_{n+1}$, and choosing $\phi_2 = \phi_{n+1}$ and $\phi_4 = \phi_n$ (and similarly for ψ_1 and ψ_3).

As they are so short, the decomposition cobweb iterative schemes analogous to schemes B and D will be developed in this section. The decomposition optimising scheme analogous to scheme B will be developed in this section, while that analogous to scheme D will be developed in Appendix VII. The schemes and their conditions for convergence will be tabulated in tables (IV.5.1) and (IV.5.2), and then the section ends with the application to an example.

Cobweb Iterative Schemes

As in the last section, we want to find composite operators P and R such that

$$\phi_{n+1} - \phi_e = P(\phi_n - \phi_e) \text{ and } \psi_{n+1} - \psi_e = R(\phi_{n+1} - \phi_e) \qquad \text{(IV.5.8)}$$

$$\text{or } \psi_{n+1} - \psi_e = P(\psi_n - \psi_e) \text{ and } \phi_{n+1} - \phi_e = R(\psi_{n+1} - \psi_e) \qquad \text{(IV.5.9)}$$

Then, by theorem (II.16.1),

$$\lim_{n \to \infty} \quad \| \phi_{n+1} - \phi_e \| + \| \psi_{n+1} - \psi_e \| = 0$$

if P is a linear, self-adjoint operator such that, for some real numbers q and Q

$$- I < q I \leq P \leq QI < I \qquad \text{(IV.5.10)}$$

B. Let $\phi_1 = \phi_2 = \phi_4 = \phi_n$, $\psi_1 = \psi_2 = \psi_n$ and $\psi_3 = \psi_{n+1}$ (IV.5.11)

then equations (IV.5.2) and (IV.5.3) become

$$A\psi_n + B\phi_n + f = 0 \qquad \text{(IV.5.12)}$$

$$\text{and } A^x\phi_n - C_M\psi_n - C_N\psi_{n+1} + g = 0 \qquad \text{(IV.5.13)}$$

From the above two equations and equations (IV.5.6) and (IV.5.7), we have

$$C_N(\psi_{n+1} - \psi_e) = A^x(\phi_n - \phi_e) - C_M(\psi_n - \psi_e) \qquad \text{(IV.5.14)}$$

$$\text{and } \phi_n - \phi_e = - B^{-1} A(\psi_n - \psi_e) \qquad \text{(IV.5.15)}$$

Substituting equation (IV.5.15) into (IV.5.14) gives

$$C_N(\psi_{n+1} - \psi_e) = - (A^x B^{-1} A + C_M)(\psi_n - \psi_e)$$

$$\text{or } \psi_{n+1} - \psi_e = - C_N^{-1}(A^x B^{-1} A + C_M)(\psi_n - \psi_e)$$

$$\text{giving } P = - C_N^{-1}(A^x B^{-1} A + C_M) \qquad \text{(IV.5.16)}$$

D. Let $\phi_1 = \phi_4 = \phi_n$, $\phi_2 = \phi_{n+1}$ and $\psi_1 = \psi_2 = \psi_3 = \psi_{n+1}$

(IV.5.17)

Equations (IV.5.2) and (IV.5.3) then become

$$A^x \phi_n - C \psi_n + g = 0 \qquad \text{(IV.5.18)}$$

and $A\psi_n + B_N \phi_n + B_M \phi_{n+1} + f = 0$ (IV.5.19)

From these two equations and equations (IV.5.6) and (IV.5.7) we have

$$B_M (\phi_{n+1} - \phi_e) = - A(\psi_n - \psi_e) - B_N (\phi_n - \phi_e) \qquad \text{(IV.5.20)}$$

and $\psi_n - \psi_e = C^{-1} A^x (\phi_n - \phi_e)$ (IV.5.21)

Substituting equation (IV.5.21) into (IV.5.20) results in

$$B_M (\phi_{n+1} - \phi_e) = - (AC^{-1} A^x + B_N) (\phi_n - \phi_e)$$

or $\phi_{n+1} - \phi_e = - B_M^{-1} (AC^{-1} A^x + B_N) (\phi_n - \phi_e)$

which gives

$$P = - B_M^{-1} (AC^{-1} A^x + B_N) \qquad \text{(IV.5.22)}$$

The results from above are summarised in the following table.

Table (IV.5.1): Decomposition Cobweb Iterative Schemes

The cobweb iterative schemes specified by the given pair of equations will converge if P is a linear, self-adjoint operator and satisfies

$$- I < q I \leq P \leq QI < I$$

where q and Q are real numbers.

B. $A\psi_n + B\phi_n + f = 0$

and $A^x \phi_n - C_M \psi_n - C_N \psi_{n+1} + g = 0$

$P = - C_N (A^x B^{-1} A + C_M)$

D. $A^x \phi_n - C\psi_n + g = 0$

and $A\phi_n + B_N \psi_n + B_M \psi_{n+1} + f = 0$

$P = - B_M^{-1} (AC^{-1} A^x + B_N)$.

Which scheme is used does, of course, depend on how easy it is to find the inverse operators in any given problem.

CHAPTER IV

Optimising Iterative Schemes

We start from equations (IV.5.1) to (IV.5.3), using the method developed in Chapter III. As an alternative to calculus, to find the required maximum or minimum we are going to use the completed-square form of the relevant quadratic functions.

B. Let $\phi_1 = \phi_2 = \phi_4 = \phi_{2n}$, $\psi_1 = \psi_2 = \psi_{2n}$ and $\psi_3 = \psi_{2n+1}$;

then the defining equations are:

$$A\,\psi_{2n} + B\phi_{2n} + f = 0 \qquad \text{(IV.5.23)}$$

$$A^x\,\phi_{2n} - C_M\psi_{2n} - C_N\,\psi_{2n+1} + g = 0 \qquad \text{(IV.5.24)}$$

$$A\,\psi_{2n+1} + B\phi_{2n+1} + f = 0 \qquad \text{(IV.5.25)}$$

$$\phi_{2n+2} = \lambda_{2n+2}\,(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1} \qquad \text{(IV.5.26)}$$

$$\psi_{2n+2} = \lambda_{2n+2}\,(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1} \qquad \text{(IV.5.27)}$$

where λ_{2n+2} is chosen to maximise the lower bound

$$L(\phi_{2n+2}, \psi_{2n+2})_{\mathcal{B}} = -\tfrac{1}{2}\langle \phi_{2n+2}, B\,\phi_{2n+2}\rangle - \tfrac{1}{2}\langle \psi_{2n+2}, C\psi_{2n+2}\rangle + \langle \psi_{2n+2}, g\rangle \qquad \text{(IV.5.28)}$$

Substituting equations (IV.5.26) and (IV.5.27) into (IV.5.28) gives

$$L(\phi_{2n+2}, \psi_{2n+2})_{\mathcal{B}} = -\frac{P}{2}\left(\lambda_{2n+2} + \frac{Q}{P}\right)^2 + R + \frac{Q^2}{2P} \qquad \text{(IV.5.29)}$$

where $P = \langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle$ (IV.5.30)

$Q = \langle \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1} - g\rangle$ (IV.5.31)

and $R = L(\phi_{2n+1}, \psi_{2n+1})_{\mathcal{B}}$ (IV.5.32)

$L(\phi_{2n+2}, \psi_{2n+2})_{\mathcal{B}}$ takes its maximum when $\lambda_{2n+2} = -\frac{Q}{P}$

Using equations (IV.5.23) to (IV.5.25), Q can be written as

$$Q = -\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle - \langle \psi_{2n} - \psi_{2n+1}, C_M(\psi_{2n} - \psi_{2n+1})\rangle$$

and hence the best λ_{2n+2} is

$$\lambda_{2n+2} = \frac{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C_M (\psi_{2n} - \psi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C (\psi_{2n} - \psi_{2n+1})\rangle}$$

$$= 1 - \frac{\langle \psi_{2n} - \psi_{2n+1}, C_N (\psi_{2n} - \psi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle} \qquad \text{(IV.5.33)}$$

We can now deduce that $L(\phi_{2n}, \psi_{2n})_\beta \leq L(\phi_{2n+2}, \psi_{2n+2})_\beta$ when the iteration specified by equations (IV.5.23) to (IV.5.27) and (IV.5.33) are carried out.

For convergence of the iteration, using theorem (II.17.1), we need to find a linear, self-adjoint, positive-definite operator P such that the equations in the iterative scheme are equivalent to

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle Z_{2n}, Z_{2n}\rangle Z_{2n}}{\langle Z_{2n}, PZ_{2n}\rangle} \qquad \text{(IV.5.34)}$$

where $Z_{2n} = P\psi_{2n} - G$ and G satisfies $P\psi_e = G$ (IV.5.35)

Then $\lim_{n\to\infty} \|\phi_{2n+2} - \phi_e\| + \|\psi_{2n+2} - \psi_e\| = 0$

From equations (IV.5.27) and (IV.5.33), the iteration equations involving ψ_{2n+2} can be written

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle \psi_{2n} - \psi_{2n+1}, C_N (\psi_{2n} - \psi_{2n+1})\rangle (\psi_{2n} - \psi_{2n+1})}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}$$

or, using equations (IV.5.23) and (IV.5.25),

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle \psi_{2n} - \psi_{2n+1}, C_N (\psi_{2n} - \psi_{2n+1})\rangle (\psi_{2n} - \psi_{2n+1})}{\langle \psi_{2n} - \psi_{2n+1}, (A^x B^{-1} A + C) (\psi_{2n} - \psi_{2n+1})\rangle} \qquad \text{(IV.5.36)}$$

If we take $C_N = \frac{1}{n^2} I$, where n is a real, non-zero number, then equation (IV.5.36) becomes

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle \psi_{2n} - \psi_{2n+1}, \psi_{2n} - \psi_{2n+1}\rangle (\psi_{2n} - \psi_{2n+1})}{\langle \psi_{2n} - \psi_{2n+1}, n^2 (A^x B^{-1} A + C) (\psi_{2n} - \psi_{2n+1})\rangle} \qquad \text{(IV.5.37)}$$

Comparing equations (IV.5.34) and (IV.5.37) we thus need

$$Z_{2n} = \psi_{2n} - \psi_{2n+1} \quad \text{and} \quad P = n^2 (A^x B^{-1} A + C) \qquad \text{(IV.5.38)}$$

From equations (IV.5.44) and (IV.5.55),

$(A^x B^{-1} A + C)\psi_e = g - A^x B^{-1} f$

or $n^2 (A^x B^{-1} A + C)\psi_e = n^2 (g - A^x B^{-1} f)$ (IV.5.39)

Then, from equation (IV.5.35), we must take

$$G = n^2 (g - A^x B^{-1} f) \qquad \text{(IV.5.40)}$$

and so, for convergence, we require that the iteration equations (IV.5.23) to (IV.5.25) are equivalent to

$$\psi_{2n} - \psi_{2n+1} = n^2 ((A^x B^{-1} A + C)\psi_{2n} + A^x B^{-1} f - g) \qquad \text{(IV.5.41)}$$

Using equations (IV.5.23) to (IV.5.25),

$$\begin{aligned} C_N (\psi_{2n} - \psi_{2n+1}) &= C_N \psi_{2n} + C_M \psi_{2n} - A^x \phi_{2n} - g \\ &= C \psi_{2n} + A^x B^{-1} A \psi_{2n} + A^x B^{-1} f - g \\ &= (A^x B^{-1} A + C)\psi_{2n} + A^x B^{-1} f - g \end{aligned}$$

as we are taking $C_N = \frac{1}{n^2} I$,

$$\psi_{2n} - \psi_{2n+1} = n^2 ((A^x B^{-1} A + C)\psi_{2n} + A^x B^{-1} f - g),$$

which is the same as equation (IV.5.41); hence the iteration given by equations (IV.5.23) to (IV.5.27) and (IV.5.33) will converge provided $C_N = \frac{1}{n^2} I$ and $P = n^2 (A^x B^{-1} A + C)$ is linear, self-adjoint and positive-definite.

The results from above and Appendix VII are summarised in the following table.

<u>Table (IV.5.2): Decomposition Optimising Iterative Schemes</u>

B. Let $\phi_1 = \phi_2 = \phi_4 = \phi_{2n}$, $\psi_1 = \psi_2 = \psi_{2n}$ and $\psi_3 = \psi_{2n+1}$

Then the iterative scheme is:

Choose (ϕ_o, ψ_o) : $A\psi_o + B\phi_o + f = 0$

Then, for $n = 0,1,2,\ldots$ find ψ_{2n+1}: $A^x \phi_{2n} - C_M \psi_{2n} - C_N \psi_{2n+1} + g = 0$

find ϕ_{2n+1}: $A\psi_{2n} + B\phi_{2n+1} + f = 0$

Find $\lambda_{2n+2} = 1 - \dfrac{\langle \psi_{2n} - \psi_{2n+1},\ C_N(\psi_{2n} - \psi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1},\ C(\psi_{2n} - \psi_{2n+1})\rangle}$

Let $\phi_{2n+2} = \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$

$\psi_{2n+2} = \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$

We have: $L(\phi_e, \psi_e) \geqslant L(\phi_{2n+2}, \psi_{2n+2})_s \geqslant L(\phi_{2n}, \psi_{2n})_s \geqslant$

$\ldots\ldots \geqslant L(\phi_o, \psi_o)$

and, if $C_N = \frac{1}{n^2} I$ where n is a real non-zero number, and $n^2 (A^x B^{-1} A + C)$

is linear, self-adjoint and positive-definite, then

$$\lim_{n \to \infty} \|\phi_{2n+2} - \phi_e\| + \|\psi_{2n+2} - \psi_e\| = 0$$

D. Let $\phi_1 = \phi_4 = \phi_{2n}$, $\phi_2 = \phi_{2n+1}$ and $\psi_1 = \psi_2 = \psi_3 = \psi_{2n}$

Then the iterative scheme is:

Choose (ϕ_o, ψ_o) : $A^x \phi_o - C\psi_o + g = 0$

Then, for n = 0,1,2,... find ϕ_{2n+1}: $A\psi_{2n} + B_N \phi_{2n} + B_M \phi_{2n+1} + f = 0$

find ψ_{2n+1}: $A^x \phi_{2n+1} - C\psi_{2n+1} + g = 0$

Find $\lambda_{2n+2} = 1 - \dfrac{\langle \phi_{2n} - \phi_{2n+1},\ B_M(\phi_{2n} - \phi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1},\ C(\psi_{2n} - \psi_{2n+1})\rangle}$

Let $\phi_{2n+2} = \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$

$\psi_{2n+2} = \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1}$

We have $L(\phi_e, \psi_e) \leq L(\phi_{2n+2}, \psi_{2n+2})_\alpha \leq L(\phi_{2n}, \psi_{2n})_\alpha$

$\leq \ldots\ldots \leq L(\phi_o, \psi_o)_\alpha$

and, if $B_M = \frac{1}{m^2} I$ where m is a real non-zero number, and $m^2 (B + AC^{-1} A^x)$

is linear, self-adjoint and positive-definite, then

$$\lim_{n \to \infty} \|\phi_{2n+2} - \phi_e\| + \|\psi_{2n+2} - \psi_e\| = 0$$

As in the cobweb iterative schemes, which of the above iterative schemes is used in a particular problem depends on the ease by which the inverse operators can be found.

From tables (IV.4.1) and (IV.5.2), it can be seen that, as in the two iterative schemes applied to classical dual extremum principles, the conditions for convergence in the cobweb iterative scheme are generally more restrictive than the convergence conditions in the optimising iterative schemes; for example, if we choose $A = A^x = B^{-1} = I$, for cobweb scheme B to converge we require

$$-I < qI \leq -C_N^{-1}(C_M + I) \leq QI < I, \quad q\ Q \in \mathbb{R}$$

but for the optimising scheme B to converge we just require that C is a bounded operator.

We end this section with an example which compares the conditions necessary for convergence using the various schemes. The conditions are found by substituting in the given operators into the conditions given in Tables (IV.4.1), (IV.4.2), (IV.5.1) and (IV.5.2). As we want to make comparisons between schemes, only schemes B and D will be considered.

<u>Example</u>

In equations (IV.5.4) and (IV.5.5), let

$$A = A^x = C = I \text{ and } g = 0 \tag{IV.5.42}$$

and also let the inverse of B be given by $B^{-1} = T$ (IV.5.43)

Then the problem under consideration is given by the equations

$$\psi_e + B\phi_e + f = 0 \text{ and } \phi_e - \psi_e = 0, \text{ or}$$

$$(B + I)\phi_e = -f \tag{IV.5.44}$$

Table (IV.5.3) details the convergence conditions for the various iterative schemes.

Table (IV.5.3): Conditions for Convergence

	Scheme	Conditions
Table (IV.4.1)	Cobweb/ Classical B Cobweb/ Classical D	$-I < qI \leq -T \leq QI < I, \quad q, Q \in \mathbb{R}$
Table (IV.4.2)	Optimising/ Classical B	T is linear and self-adjoint and T + I is positive-definite
	Optimising/ Classical D	B = I gives $2\phi_e = -f$, which is not a useful problem!
Table (IV.5.1)	Cobweb/ Decomposition B	$-I < qI \leq -C_N^{-1}(T + C_M) \leq QI < I, \quad q, Q \in \mathbb{R}$ where $C_N + C_M = I$
	Cobweb/ Decomposition D	$-I < qI \leq -B_M^{-1}(I + B_N) \leq QI < I, \quad q, Q \in \mathbb{R}$ where $B_M + B_N = B$
Table (IV.5.2)	Optimising/ Decomposition B	$C_N = \frac{1}{n^2} I$ and T + I is linear, self-adjoint and positive-definite; $C_M + C_N = I$
	Optimising/ Decomposition D	$B_M = \frac{1}{m^2} I$ and $(B + I)$ is linear, self-adjoint and positive-definite; $B_M + B_N = B$

Obviously, whether the convergence conditions in Table (IV.5.3) can be satisfied depends on the structure of operator B. We are going to assume that one of B and T is a linear, positive-definite symmetric differential operator, which is not bounded above, and the other, its inverse is the corresponding linear, positive-definite, self-adjoint integral operator, which is bounded.

(a) Let B be a differential operator and T the corresponding integral operator. From Table (IV.5.3) we can work out for which iterative schemes convergence is possible.

<u>Cobweb/Classical B,D</u>

As T is positive-definite and symmetric, we require $T \leq -qI < I$.

Optimising/Classical B

No further conditions on T necessary.

Optimising Classical D

Not applicable.

Cobweb/Decomposition B

Let $C_N = \frac{1}{n^2} I$, then $C_M = (1 - \frac{1}{n^2})I$, and hence for convergence we require

$$0 < T \leq \left(\frac{1 - q - n^2}{n^2}\right)I < \left(\frac{2}{n^2} - 1\right)I$$

Cobweb/Decomposition D

Not applicable, as we would require $B_N = B - B_M$ to be bounded.

Optimising/Decomposition B

No further conditions necessary.

Optimising/Decomposition D

Not applicable, as we would require B to be bounded.

From the above, it is obvious that optimising schemes B, both classical and decomposition, are better than the cobweb schemes, as no further restrictions on T are necessary; furthermore, there is no advantage in using the decomposition bounds for the optimising iterative scheme. However, it is sometimes easier to use the cobweb iterative schemes (see section III.11), and whether we use the classical or decomposition cobweb iterative schemes depends on T.

If $T < I$, then we can use cobweb/classical scheme B or D. If there exists a finite, non-zero parameter n such that $I \leq T < (\frac{2}{n^2} - 1)I$, then cobweb/decomposition scheme B could be used.

(b) Let B be an integral operator and T the corresponding differential operator. From table (IV.3.3), the iterative schemes which converge can be found.

As T is an unbounded operator, none of the classical iterative schemes can be shown to converge, and neither can the cobweb or optimising decomposition schemes β.

Cobweb/Decomposition D

Let $B_M = \frac{1}{m^2} I$ giving $B_N = B - \frac{1}{m^2} I$. Then this scheme will converge provided B satisfies

$$0 < B \leq \left(\frac{1 - q - m^2}{m^2}\right) I < \left(\frac{2}{m^2} - 1\right) I \quad \text{(m not equal to zero).}$$

Optimising/Decomposition D

No further conditions are necessary on B.

Obviously the optimising decomposition iteration D is the best to use, but if $B < \left(\frac{2}{m^2} - 1\right)I$, cobweb/decomposition iteration D could be used, especially if it is easier to use than the corresponding optimising scheme.

- Of course, the fact that we cannot prove convergence for some types of problems does not mean that we cannot use iterative schemes for these problems, and it is still possible that convergence could occur either to (ϕ_e, ψ_e) or $L(\phi_e, \psi_e)$.

CHAPTER IV

IV.6 Applications

In this section, we are going to apply the decomposition of functionals theorem to four particular examples.

The first example shows how pointwise bounds can be obtained for the solution of an integral equation, using the methods in this chapter instead of the bivariational bounds method developed by Robinson and Barnsley.

Example 2 considers a new application, see Bailey and Smith, (15) of dual extremum principles to periodic solutions of a group of problems defined by the equation

$\frac{d^2\phi}{dt^2} + (K - \gamma^2)\phi + g'(\phi) = \Gamma \sin w t$, where ϕ is a real function which is $\frac{2\pi}{w}$ - periodic and satisfies $\phi(0) = 0$. K, γ, Γ, and w are real numbers and $g(\phi)$ is an even, concave function.

The third example sets up an iterative scheme for a general functional $L(\phi, \psi)$, and considers conditions for convergence of the scheme to $L(\phi_e, \psi_e)$.

The last example considers the boundary value problem $\nabla^2 u + F'(u) = 0$ in V, $u = u_S$ on S, and uses a new functional to obtain the flexibility required to apply theorem (IV.2.2).

CHAPTER IV

Example 1

In the survey, in section I.8, a review was made of the papers on Bivariational bounds. Paper (55) considered the equation $A\phi = f$, $f \in \mathcal{H}$ where A is a linear, self-adjoint, positive-definite operator, and found complementary bivariational bounds to the quantity $\langle \phi, g \rangle$ where $g \in \mathcal{H}$. The example discussed in paper (55) was the integral equation

$$\phi(x) + \lambda \int_a^b k(x,y)\phi(y)\, dy = f(x); \qquad \text{(IV.6.1)}$$

by choosing g as k(x,y), pointwise bounds were obtained on $\phi(y)$, as

$$\langle \phi, g \rangle = \frac{1}{\lambda}(f(y) - \phi(y)) \qquad \text{(IV.6.2)}$$

In this section we are going to show how different pointwise bounds can be obtained by the use of the decomposition of functionals theorem, theorem (IV.2.2).

We define the functional $L(\phi, \psi) : E^1 \times E^1 \to \mathbb{R}$ by the equation

$$L(\phi, \psi) = \tfrac{1}{2}\langle \phi, B\phi \rangle - \tfrac{1}{2}\langle \psi, B\psi \rangle + \langle f, \phi \rangle + \langle g, \psi \rangle \qquad \text{(IV.6.3)}$$

where B is a linear, symmetric, positive-definite operator and f and g are functions; then $L(\phi, \psi)$ is a convex/concave saddle functional. The functional derivatives of $L(\phi, \psi)$ are given by

$$\nabla_\phi L(\phi, \psi) = B\phi + f \qquad \text{(IV.6.4)}$$

and $$\nabla_\psi L(\phi, \psi) = -B\psi + g \qquad \text{(IV.6.5)}$$

The stationary point occurs at (ϕ_e, ψ_e), where

$$B\phi_e + f = 0 \qquad \text{(IV.6.6)}$$

and $$-B\psi_e + g = 0 \qquad \text{(IV.6.7)}$$

and the stationary value $L(\phi_e, \psi_e)$ is given by

$$L(\phi_e, \psi_e) = \tfrac{1}{2}\langle \phi_e, f \rangle + \tfrac{1}{2}\langle \psi_e, g \rangle \qquad \text{(IV.6.8)}$$

Adding and subtracting equations (IV.6.6) and (IV.6.7) results in

$$B(\phi_e - \psi_e) + f + g = 0 \qquad \text{(IV.6.9)}$$

and $$B(\phi_e + \psi_e) + f - g = 0 \qquad \text{(IV.6.10)}$$

Let $U_e = \phi_e - \psi_e$, $V_e = \phi_e + \psi_e$, $F = f + g$ and $G = f - g$ (IV.6.11)

Then equations (IV.6.9) and (IV.6.10) can be written

$$B\,U_e + F = 0 \quad \text{(IV.6.12)}$$

and $$B\,V_e + G = 0 \quad \text{(IV.6.13)}$$

$$\begin{aligned} \langle f, \psi_e \rangle &= -\langle B\phi_e, \psi_e \rangle \quad \text{from equation (IV.6.6)} \\ &= -\langle \phi_e, B\psi_e \rangle \\ &= -\langle \phi_e, g \rangle \quad \text{from equation (IV.6.7)} \end{aligned} \quad \text{(IV.6.14)}$$

We can therefore rewrite equation (IV.6.8) as

$$\begin{aligned} L(\phi_e, \psi_e) &= \tfrac{1}{2}\langle \phi_e, f \rangle + \tfrac{1}{2}\langle \psi_e, g \rangle - \tfrac{1}{2}\langle f, \psi_e \rangle - \tfrac{1}{2}\langle \phi_e, g \rangle \\ &= \tfrac{1}{2}\langle \phi_e - \psi_e, f - g \rangle \\ &= \tfrac{1}{2}\langle U_e, G \rangle \end{aligned} \quad \text{(IV.6.15)}$$

We now apply the decomposition of functionals theorem to the problem, theorem (IV.2.2).

Let $M(\phi, \psi) : E^1 \times E^1 \rightarrow \mathbb{R}$ be defined by the equation

$$M(\phi, \psi) = \tfrac{1}{2}\mu\langle \phi, \phi \rangle - \tfrac{1}{2}\mu\langle \psi, \psi \rangle \quad \text{(IV.6.16)}$$

where μ is a real number greater than zero.

Then, as we define $N(\phi, \psi)$ by the equation $L(\phi, \psi) = M(\phi, \psi) + N(\phi, \psi)$, we must have $N(\phi, \psi) = L(\phi, \psi) - M(\phi, \psi)$

or $$N(\phi, \psi) = \tfrac{1}{2}\langle \phi, (B - \mu I)\phi \rangle - \tfrac{1}{2}\langle \psi, (B - \mu I)\psi \rangle + \langle \phi, f \rangle + \langle \psi, g \rangle \quad \text{(IV.6.17)}$$

$M(\phi, \psi)$ is a strict convex/concave saddle functional as μ is greater than zero; then $N(\phi, \psi)$ is a convex/concave saddle functional if $B - \mu I$ is a positive operator. (IV.6.18)

The functional derivatives of $M(\phi, \psi)$ and $N(\phi, \psi)$ are

$$\nabla_\phi M(\phi, \psi) = \mu\phi \quad \text{(IV.6.19)}$$

$$\nabla_\psi M(\phi, \psi) = -\mu\psi \quad \text{(IV.6.20)}$$

$$\nabla_\phi N(\phi, \psi) = (B - \mu I)\phi + f \quad \text{(IV.6.21)}$$

$$\nabla_\psi N(\phi, \psi) = -(B - \mu I)\psi + g \quad \text{(IV.6.22)}$$

The application of theorem (IV.2.2) results in

$$\tfrac{1}{2}\mu\langle\phi_2,\phi_2\rangle + \tfrac{1}{2}\langle\phi_4, (B-\mu I)\phi_4\rangle - \tfrac{1}{2}\langle\psi_2, (B-\mu I)\psi_2\rangle$$
$$+ \langle\phi_4, f\rangle + \langle\psi_2, g\rangle - \langle\phi_2 - \phi_4, \mu\phi_2\rangle - \tfrac{1}{2}\mu\langle\psi_2,\psi_2\rangle$$
$$\leq L(\phi_e, \psi_e) \leq$$
$$\tfrac{1}{2}\mu\langle\phi_1,\phi_1\rangle - \tfrac{1}{2}\mu\langle\psi_1,\psi_1\rangle + \tfrac{1}{2}\langle\phi_1, (B-\mu I)\phi_1\rangle - \tfrac{1}{2}\langle\psi_3, (B-\mu I)\psi_3\rangle$$
$$+ \langle\phi_1, f\rangle + \langle\psi_3, g\rangle - \langle\psi_1 - \psi_3, \mu\psi_1\rangle \qquad \text{(IV.6.23)}$$

where $\mu\phi_2 + (B-\mu I)\phi_4 + f = 0$ (IV.6.24)

and $-\mu\psi_1 - (B-\mu I)\psi_3 + g = 0$ (IV.6.25)

Substituting equations (IV.6.24) and (IV.6.25) into (IV.6.23) gives

$$-\tfrac{1}{2}\mu\langle\phi_2,\phi_2\rangle - \tfrac{1}{2}\langle\psi_2, B\psi_2\rangle + \langle\psi_2, g\rangle - \tfrac{1}{2}\langle\phi_4, (B-\mu I)\phi_4\rangle$$
$$\leq L(\phi_e, \psi_e) \leq$$
$$\tfrac{1}{2}\mu\langle\phi_1, B\phi_1\rangle + \tfrac{1}{2}\mu\langle\psi_1, \psi_1\rangle + \langle\phi_1, f\rangle + \tfrac{1}{2}\langle\psi_3, (B-\mu I)\psi_3\rangle \qquad \text{(IV.6.26)}$$

Equation (IV.6.26) can be further simplified if we eliminate ϕ_2 and ψ_1 using equations (IV.6.24) and (IV.6.25). Then

$$-\tfrac{1}{2\mu}\,\| B\phi_4 + f \|^2 + \tfrac{1}{2}\langle\phi_4, B\phi_4\rangle - \tfrac{1}{2}\langle\psi_2, B\psi_2\rangle + \langle\phi_4, f\rangle + \langle\psi_2, g\rangle$$
$$\leq L(\phi_e, \psi_e) \leq$$
$$\tfrac{1}{2\mu}\,\| B\psi_3 - g \|^2 + \tfrac{1}{2}\langle\phi_1, B\phi_1\rangle - \tfrac{1}{2}\langle\psi_3, B\psi_3\rangle + \langle\phi_1, f\rangle + \langle\psi_3, g\rangle \qquad \text{(IV.6.27)}$$

Finally, if we let $\phi_1 = \phi_4 = \phi_o$, $\psi_2 = \psi_3 = \psi_o$ and use equation (IV.6.15) for $L(\phi_e, \psi_e)$, we have the bounds

$$-\tfrac{1}{2\mu}\,\| B\phi_o + f \|^2 \leq \tfrac{1}{2}\langle U_e, G\rangle - \tfrac{1}{2}\langle\phi_o, B\phi_o\rangle + \tfrac{1}{2}\langle\psi_o, B\psi_o\rangle$$
$$- \langle\phi_o, f\rangle - \langle\psi_o, g\rangle \leq \tfrac{1}{2\mu}\,\| B\psi_o - g \|^2 \qquad \text{(IV.6.28)}$$

We now consider the integral equation

$$B\,U_e(x) = (I + \lambda K)\,U_e(x) = -F(x) \qquad \text{(IV.6.29)}$$

where K, symmetric, positive-definite and self-adjoint, is defined by

$K\,U_e(x) = \int_a^b k(x,y)\,U_e(y)\,dy$, and $k(x,y)$ is a symmetric kernel. $F(x)$ is a given function continuous on $[a,b]$ and the parameter λ takes regular values to ensure that $U_e(x)$ is unique. If $G(x)$ is an arbitrary continuous function, the stationary value becomes

$$\tfrac{1}{2}\langle U_e, G\rangle = \tfrac{1}{2}\int_a^b U_e(x)\, G(x)\, dx \tag{IV.6.30}$$

In order to apply the results obtained above we require $V_e(x)$ to satisfy the dual equation

$$B\, V_e(x) = (I + \lambda K)\, V_e(x) = -\, G(x) \tag{IV.6.31}$$

We also need to be able to find μ such that $(I + \lambda K - \mu I) \geq 0$. As K is positive-definite, $K > 0$; then $(I + \lambda K - \mu I) > (1 - \mu)I$ which is positive if $\mu \leq 1$.

The bounds given by equation (IV.6.28) become for this integral equation

$$-\frac{1}{2\mu}\int_a^b \left(\phi_o(x) + \lambda\int_a^b k(x,y)\,\phi_o(y)\,dy + f(x)\right)^2 dx$$

$$\leq \tfrac{1}{2}\int_a^b U_e(x)\, G(x)\, dx - \tfrac{1}{2}\int_a^b \Big((\phi_o(x))^2 - (\psi_o(x))^2$$

$$+ 2\,\phi_o(x)\, f(x) + 2\psi_o(x)\, g(x) + \lambda\phi_o(x)\int_a^b k(x,y)\,\phi_o(y)\, dy$$

$$+ \lambda\psi_o(x)\int_a^b k(x,y)\,\psi_o(y)\, dy\Big)\, dx$$

$$\leq \frac{1}{2\mu}\int_a^b \left(\psi_o(x) + \lambda\int_a^b k(x,y)\,\psi_o(y)\, dy - g(x)\right)^2 dx \tag{IV.6.32}$$

Pointwise bounds for the solution of equation (IV.6.29) can be constructed by letting $G(x) = k(x,z)$ and treating z as a parameter. Then we have

$$\langle G, U_e\rangle = \int_a^b k(z,x)\, U_e(x)\, dx$$

$$= -\frac{1}{\lambda}\left(U_e(z) + F(z)\right) \tag{IV.6.33}$$

from equation (IV.6.29).

From equations (IV.6.11) and noting that $G(x) = k(x,z)$, we have

$$f(x) = \tfrac{1}{2}\left(F(x) + k(x,z)\right) \tag{IV.6.34}$$

and

$$g(x) = \tfrac{1}{2}\left(F(x) - k(x,z)\right) \tag{IV.6.35}$$

In order to compare the bounds given by equation (IV.6.32) with those given for a particular integral equation in paper (55), we are going to let $\phi_o(x) = \psi_o(x) = 0$; then equation (IV.6.32) becomes

$$-\frac{\lambda}{8\mu}\int_a^b (F(x) - k(x,z))^2\,dx \leqslant U_e(z) + F(z)$$

$$\leq \frac{\lambda}{8\mu}\int_a^b (F(x) + k(x,z))^2\,dx \qquad \text{(IV.6.36)}$$

We are going to consider the example specified by

$$a = 0, \quad b = 1, \quad \lambda = 1 \qquad \text{(IV.6.37)}$$

$$F(x) = -x^2 \text{ and } k(x,z) = \begin{cases} x\,(1-z), & x \leq z \\ z\,(1-x), & x \geqslant z \end{cases} \qquad \text{(IV.6.38)}$$

The sharpest bounds will be obtained if we take $\mu = 1$. Substituting these values and functions into equation (IV.6.36) results in the bounds

$$-\frac{1}{8}\left\{\int_o^z (x^2 + z(1-x))^2\,dx + \int_z^1 (x^2 + x(1-z))^2\,dx\right\} + z^2$$

$$\leq U_e(z) \leq$$

$$\frac{1}{8}\left\{\int_o^z (-x^2 + z(1-x))^2\,dx + \int_z^1 (-x^2 + x(1-z))^2\,dx\right\} + z^2 \qquad \text{(IV.6.39)}$$

Carrying out the integration gives

$$-\frac{1}{8}\left\{\frac{31}{30} - \frac{5z}{6} + \frac{z^2}{3} + \frac{2z^3}{3} - \frac{z^4}{6}\right\} + z^2$$

$$\leq U_e(z) \leq$$

$$\frac{1}{8}\left\{\frac{1}{30} - \frac{z}{6} + \frac{z^2}{3} + \frac{2z^3}{3} - \frac{z^4}{2}\right\} + z^2 \qquad \text{(IV.6.40)}$$

At $z = \frac{1}{2}$ we have the bounds

$$\frac{589}{3840} \leq U_e(\tfrac{1}{2}) \leq \frac{1001}{3840}$$

or $0.153385416 \leqslant U_e(\frac{1}{2}) \leqslant 0.260677083$

These are not as accurate as the bounds obtained in paper (55), but trial functions ϕ_o and ψ_o could be chosen to give sharp bounds.

CHAPTER IV

Example 2

The decomposition of functionals theorem will now be applied to a new group of problems in ordinary differential equations.

Let $L(\phi, \psi) : E^1 \times E^1 \to \mathbb{R}$ be defined by the equation

$$L(\phi, \psi) = \int_0^{\frac{2\pi}{w}} \left\{ \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 - \frac{1}{2}K\phi^2 - \frac{1}{2}\psi^2 - \gamma\phi\psi - g(\phi) + \phi\Gamma \sin wt \right\} dt \qquad \text{(IV.6.41)}$$

where K, γ, Γ and w are real numbers and $g(\phi)$ is a real differentiable function. We shall assume that ϕ and ψ belong to the space E^1 of twice differentiable real functions which are $2\pi/w$ - periodic and such that

$$\phi(0) = \psi(0) = 0 \qquad \text{(IV.6.42)}$$

The functional derivatives of $L(\phi, \psi)$ are given by

$$\nabla_\phi L(\phi, \psi) = -\frac{d^2\phi}{dt^2} - K\phi - \gamma\psi - g'(\phi) + \Gamma \sin wt \qquad \text{(IV.6.43)}$$

and $$\nabla_\psi L(\phi, \psi) = -\psi - \gamma\phi \qquad \text{(IV.6.44)}$$

At the stationary value of $L(\phi, \psi)$, the equations $\nabla_\psi L(\phi_e, \psi_e) = 0$ and $\nabla_\phi L(\phi_e, \psi_e) = 0$ imply that ϕ satisfies

$$\frac{d^2\phi_e}{dt^2} + (K - \gamma^2)\phi_e + g'(\phi_e) = \Gamma \sin wt \qquad \text{(IV.6.45)}$$

provided that there exists at least one periodic solution of this equation. The existance of at least one periodic solution is guaranteed (56), pg 194 if $g(\phi)$ satisfies the following conditions

(i) there exists a real number $M > 0$ such that, for $t \in \mathbb{R}$,

$$\Gamma \sin wt - g'(-M) + (K - \gamma^2) M < 0 \qquad \text{(IV.6.46)}$$

and

$$\Gamma \sin wt - g'(M) - (K - \gamma^2) M > 0 \qquad \text{(IV.6.47)}$$

(ii) for every $t \in \mathbb{R}$ and $\phi \in (-M, M)$,

$\Gamma \sin wt - g'(\phi) - (K - \gamma^2)$

is bounded.

In addition we require $\phi(0) = 0$; this can be proved by symmetry arguments if we assume that $g(\phi)$ is an even function of ϕ :

Let $\phi(t)$ be the solution of equation (IV.6.45)

$$\frac{d^2\phi(t)}{dt^2} + \Omega^2\phi(t) + \frac{d}{dx}g(\phi) = \Gamma \sin wt \qquad (IV.6.48)$$

then

$$\frac{d^2}{dt^2}(-\phi(t)) + \Omega^2(-\phi(t)) + \frac{d}{dx} - g(-\phi) = -\Gamma \sin wt$$

or

$$\frac{2}{t^2}(-\phi(-t)) + \Omega^2(-\phi(-t)) - \frac{d}{dx}g(\phi) = -\Gamma \sin wt \qquad (IV.6.49)$$

since $g(\phi) = g(-\phi)$.

Therefore $-\phi(-t)$ also satisfies equation (IV.6.48).

Hence $\phi(t)$ is an odd function, and as it is continuous, $\phi(0) = -\phi(-0) = 0$.

Using equation (II.9.1), $L(\phi, \psi)$ will be a convex/concave saddle functional if

$$\int_0^{\frac{2\pi}{w}} \Big\{ \tfrac{1}{2}\Big(\frac{d}{dt}(\phi_1 - \phi_2)\Big)^2 - \tfrac{1}{2}K(\phi_1 - \phi_2)^2 + \tfrac{1}{2}(\psi_1 - \psi_2)^2 + g(\phi_2) - g(\phi_1) - (\phi_2 - \phi_1)\, g'(\phi_2)\Big\}\, dt \geq 0$$

Now, $\int_0^{\frac{2\pi}{w}} \frac{1}{2}(\psi_1 - \psi_2)^2\, dt \geq 0 \quad \forall\ (\psi_1, \psi_2) \in E^1$ and, by equation (II.8.5),

$\int_0^{\frac{2\pi}{w}} (g(\phi_2) - g(\phi_1) - (\phi_2 - \phi_1)\, g'(\phi_2))\, dt \geq 0$ if $g(\phi)$ is a concave function. Hence $L(\phi, \psi)$ will be a convex/concave saddle functional if

$$\tfrac{1}{2}\int_0^{\frac{2\pi}{w}} \Big\{\frac{d}{dt}(\phi_1 - \phi_2)^2 - K(\phi_1 - \phi_2)^2\, dt\Big\} \geq 0$$

By lemma (II.3.2), assuming that $\phi(t)$ is a continuous function with a continuous derivative for $t \in (0, \frac{2\pi}{w})$, we have

$$\frac{1}{2}\int_0^{\frac{2\pi}{w}} \left(\frac{d}{dt}(\phi_1 - \phi_2)\right)^2 dt \geq \frac{w^2}{4\pi^2}\int_0^{\frac{2\pi}{w}} (\phi_1 - \phi_2)^2 \, dt$$

and hence

$$\frac{1}{2}\int_0^{\frac{2\pi}{w}} \left\{\left(\frac{d}{dt}(\phi_1 - \phi_2)\right)^2 - K(\phi_1 - \phi_2)^2\right\} dt$$

$$\geq \frac{1}{2}\left(\frac{w^2}{2\pi^2} - K\right)\int_0^{\frac{2\pi}{w}} (\phi_1 - \phi_2)^2 \, dt; \qquad \text{(IV.6.50)}$$

therefore $L(\phi, \psi)$ is a convex/concave saddle functional if

$$K \leq \frac{w^2}{2\pi^2} \text{ and } g(\phi) \text{ is a concave function.} \qquad \text{(IV.6.51)}$$

Interest usually centres on the equation in which $K > \gamma^2$, thus we shall assume that $\frac{w^2}{2\pi^2} \geq K \geq \gamma^2$, and put $K - \gamma^2 = \Omega^2$. Bounds will be found for the particular equation with $g(\phi) = -\frac{1}{4}\varepsilon\phi^4$, where $\varepsilon > 0$. In this case we are interested in periodic solutions of the equation

$$\frac{d^2\phi}{dt^2} + \Omega^2\phi - \varepsilon\phi^3 = \Gamma \sin wt \qquad \text{(IV.6.52)}$$

This is the undamped Duffing equation with harmonic forcing. It is shown in (56), pg 195 that for this choice of $g(\phi)$ the conditions given earlier are satisfied and so there exists at least one periodic solution of equation (IV.6.52).

If ϕ_e is an exact solution of equation (IV.6.52) then the stationary value of $L(\phi, \psi)$ becomes

$$L(\phi_e, \psi_e) = \int_0^{\frac{2\pi}{w}} \left\{\frac{1}{2}\phi_e \Gamma \sin wt - \frac{1}{4}\varepsilon\phi_e^4\right\} dt \qquad \text{(IV.6.53)}$$

which is not in itself a number of particular physical interest; but close upper and lower bounds imply broad global properties of trial solutions which produce the bounds. The classical dual extremum principles for the problem are $L(\phi_\sigma, \psi_\sigma) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha)$, where

$$L(\phi_\beta, \psi_\beta) = \int_0^{\frac{2\pi}{w}} \left\{ \tfrac{1}{2}\phi_\beta \frac{d^2\phi_\beta}{dt^2} + \tfrac{1}{2}K\phi_\beta^2 - \tfrac{1}{2}\psi_\beta^2 - \tfrac{1}{4}\epsilon\,\phi_\beta^4 \right\} dt \qquad \text{(IV.6.54)}$$

with $\frac{d^2\phi_\beta}{dt^2} + K\phi_\beta + \gamma\psi_\beta - \epsilon\phi_\beta^3 - \Gamma\sin wt = 0$

$$\phi_\beta(t) = \phi_\beta(t + \tfrac{2\pi}{w}),\quad \psi_\beta(t) = \psi_\beta(t + \tfrac{2\pi}{w}) \text{ and } \phi_\beta(0) = \psi_\beta(0) = 0 \qquad \text{(IV.6.55)}$$

and

$$L(\phi_\alpha, \psi_\alpha) = \int_0^{\frac{2\pi}{w}} \left\{ \tfrac{1}{2}\left(\frac{d\phi_\alpha}{dt}\right)^2 - \tfrac{1}{2}\Omega^2\phi_\alpha^2 + \tfrac{1}{4}\epsilon\,\phi_\alpha^4 + \Gamma\phi_\alpha \sin wt \right\} dt \qquad \text{(IV.6.56)}$$

where $\phi_\alpha(t) = \phi_\alpha(t + \frac{2\pi}{w})$, and $\phi_\alpha(0) = 0.$ (IV.6.57)

To apply the decomposition of functionals theorem, theorem (IV.2.2), to the problem, we let $M(\phi, \psi)$ be defined by the equation

$$M(\phi, \psi) = \int_0^{\frac{2\pi}{w}} \left\{ \tfrac{1}{2}\left(\frac{d\phi}{dt}\right)^2 - \tfrac{1}{2}K\phi^2 \right\} dt, \qquad \text{(IV.6.58)}$$

so that $$N(\phi, \psi) = \int_0^{\frac{2\pi}{w}} \left\{ -\tfrac{1}{2}\psi^2 - \gamma\phi\psi + \tfrac{1}{4}\epsilon\,\phi^4 + \Gamma\phi\sin wt \right\} dt \qquad \text{(IV.6.59)}$$

$M(\phi, \psi)$ and $N(\phi, \psi)$ are both convex/concave saddle functionals such that $L(\phi, \psi) = M(\phi, \psi) + N(\phi, \psi)$. The functional derivatives of $M(\phi, \psi)$ and $N(\phi, \psi)$ are given by the equations

$$\nabla_\phi\, M(\phi, \psi) : -\frac{d^2\phi}{dt^2} - K\phi \qquad \text{(IV.6.60)}$$

$$\nabla_\psi\, M(\phi, \psi) = 0 \qquad \text{(IV.6.61)}$$

$$\nabla_\phi\, N(\phi, \psi) = -\gamma\psi + \epsilon\phi^3 + \Gamma\sin wt \qquad \text{(IV.6.62)}$$

$$\nabla_\psi\, N(\phi, \psi) = -\psi - \gamma\phi \qquad \text{(IV.6.63)}$$

Applying theorem (IV.2.2) results in

$$\int_0^{\frac{2\pi}{w}} \left\{ \tfrac{1}{2}\left(\frac{d\phi_2}{dt}\right)^2 - \tfrac{1}{2}K\phi_2^2 - \tfrac{1}{2}\psi_2^2 - \gamma\phi_4\psi_2 + \tfrac{1}{4}\epsilon\,\phi_4^4 + \Gamma\phi_4\sin wt - (\phi_2 - \phi_4)\left(-\frac{d^2\phi_2}{dt^2} - K\phi_2\right) \right\} dt$$

$$\leq L(\phi_e, \psi_e) \leq$$

CHAPTER IV

$$\int_0^{\frac{2\pi}{w}} \left\{ \frac{1}{2}\left(\frac{d\phi_1}{dt}\right)^2 - \frac{1}{2}K\phi_1^2 - \frac{1}{2}\psi_3^2 - \gamma\phi_1\psi_3 + \frac{1}{4}\varepsilon\phi_1^4 + \Gamma \sin wt \right\} dt \tag{IV.6.64}$$

where $-\frac{d^2\phi_2}{dt^2} - K\phi_2 - \gamma\psi_2 + \varepsilon\phi_4^3 + \Gamma \sin wt = 0$ (IV.6.65)

and $-\psi_3 - \gamma\phi_1 = 0$ (IV.6.66)

Using the last two equations, equation (IV.6.64) becomes

$$\int_0^{\frac{2\pi}{w}} \left\{ \frac{1}{2}\phi_2 \frac{d^2\phi_2}{dt^2} + \frac{1}{2}K\phi_2^2 - \frac{1}{2}\psi_2^2 - \frac{3}{4}\varepsilon\phi_4^4 \right\} dt$$

$$\leq L(\phi_e, \psi_e) \leq$$

$$\int_0^{\frac{2\pi}{w}} \left\{ \frac{1}{2}\left(\frac{d\phi_1}{dt}\right)^2 - \frac{1}{2}\Omega^2\phi_1^2 + \frac{1}{4}\varepsilon\phi_1^4 + \Gamma\phi_1 \sin wt \right\} dt \tag{IV.6.67}$$

where ϕ_1, ϕ_2, ϕ_4 and ψ_2 are $\frac{2\pi}{w}$ - periodic and vanish at $t = 0$.

It is obvious, on comparing the decomposition upper bound given in equation (IV.6.67) with the classical upper bound given in equation (IV.6.56) that they are the same; this could have been predicted from the fact that $M(\phi, \psi)$ does not contain any terms in ψ. The decomposition lower bound is different from the classical lower bound as the lower bound has three trial functions in the decomposition bound compared with two trial functions in the classical bound; this increases the flexibility for the choice of trial functions in the decomposition lower bound.

We are now going to obtain bounds using equations (IV.6.65) and (IV.6.67). Let $\phi_1 = a \sin wt$, where a is a parameter to be determined. Inserting this into the upper bound in equation (IV.6.67), and calling this upper bound L_D, gives

$$L_D = -\frac{\pi a}{16w}\left(8a(\Omega^2 - w^2) - 16\Gamma - 3\varepsilon a^3\right) \tag{IV.6.68}$$

This bound takes its minimum value when $\frac{\partial L_D}{\partial a} = 0$, that is, when the amplitude a satisfies the cubic equation

$$3\varepsilon a^3 - 4a(\Omega^2 - w^2) + 4\Gamma = 0 \tag{IV.6.69}$$

This equation displays the well-known frequency-amplitude relation obtained by perturbation and averaging methods (see (41), ch 5 for a general account of the forced oscillations of the Duffing oscillator). The cubic has three real roots for the amplitude if $\Omega^2 > w^2$ and one real root if $\Omega^2 < w^2$. Now, $\frac{w^2}{2\pi^2} \geq K = \gamma^2 + \Omega^2$; therefore $\Omega^2 \leq \frac{w^2}{2\pi^2} - \gamma^2 \leq \frac{w^2}{2\pi^2} << w^2$, and hence $\Omega^2 < w^2$ and there is only one real root of the cubic equation.

Let $\phi_4 = 0$ and $\psi_2 = -\gamma b \sin wt$ (IV.6.70)

Then we require ϕ_2 to satisfy

$$\left(\frac{d^2}{dt^2} + K\right)\phi_2 = (\Gamma + \gamma^2 b) \sin wt \tag{IV.6.71}$$

A particular solution of this equation is

$$\phi_2 = \frac{\Gamma + b\gamma^2}{K - w^2} \sin wt \tag{IV.6.72}$$

Substituting these trial functions into the lower bound in equation (IV.6.67), and calling this lower bound L_D^1, gives

$$L_D^1 = -\tfrac{1}{2}\,\frac{\pi}{w}\left\{\gamma^2 b^2 - \frac{(\Gamma + b\gamma^2)^2}{(K - w^2)}\right\} \tag{IV.6.73}$$

This bound takes its largest value when $\frac{\partial L_D^1}{\partial b} = 0$, that is when $b = \frac{\Gamma}{\Omega^2 - w^2}$

(IV.6.74)

Substituting this value for b into equation (IV.6.74) gives

$$L_D^1 = -\frac{\Gamma^2\pi}{2w\,(w^2 - \Omega^2)} \tag{IV.6.75}$$

We then have $L_D - L_D^1 =$

$$\frac{\Gamma^2\pi}{2w\,(w^2 - \Omega^2)} - \frac{\pi a}{16w}\left(8a(\Omega^2 - w^2) - 16\Gamma - 3\epsilon a^3\right) \tag{IV.6.76}$$

where a is given by equation (IV.6.69).

Finally, we find a measure of the error given by equation (IV.6.76) for $0 < \epsilon < 1$.

Let $a = a_o + \epsilon a_1$, where a_o and a_1 are the first two terms in the expansion of a in powers of ϵ.

Substituting this into equation (IV.6.69) gives

$$a_0 = \frac{\Gamma}{(\Omega^2 - w^2)} \quad \text{and } a_1 = \frac{3\Gamma^3}{4(\Omega^2 - w^2)^4} \tag{IV.6.77}$$

Substituting these values into equation (IV.6.76) gives

$$L_D - L_D^1 = \frac{3\pi \Gamma^4 \varepsilon}{16w\ (\Omega^2 - w^2)^4} + 0\ (\varepsilon^2) \tag{IV.6.78}$$

(which seems to imply that $L_D - L_D^1 = 0$ if $\varepsilon = 0$). Obviously this error could be reduced if a non-zero trial function ϕ_4 was used in the lower bound in equation (IV.6.67).

Example 3

Let $L(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ be a convex/concave saddle functional. To apply theorem (IV.2.2) we require convex/concave saddle functionals $M(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ and $N(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ such that $L(\phi,\psi) = M(\phi,\psi) + N(\phi,\psi)$.

Let $M(\phi,\psi) = \frac{1}{2}\delta<\phi,\phi> - \frac{1}{2}\mu<\psi,\psi>$ (IV.6.79)

where δ and μ are real numbers greater than zero; then $M(\phi,\psi)$ is a strict convex/concave saddle functional.

We then must have $N(\phi,\psi)$ given by the equation

$$N(\phi,\psi) = L(\phi,\psi) - \tfrac{1}{2}\delta<\phi,\phi> + \tfrac{1}{2}\mu<\psi,\psi> \qquad \text{(IV.6.80)}$$

δ and μ must have values consistant with $N(\phi,\psi)$ being a convex/concave saddle functional. The functional derivatives of $M(\phi,\psi)$ and $N(\phi,\psi)$ are

$$\nabla_\phi M(\phi,\psi) = \delta\phi \qquad \text{(IV.6.81)}$$

$$\nabla_\psi M(\phi,\psi) = -\mu\psi \qquad \text{(IV.6.82)}$$

$$\nabla_\phi N(\phi,\psi) = \nabla_\phi L(\phi,\psi) - \delta\phi \qquad \text{(IV.6.83)}$$

$$\nabla_\psi N(\phi,\psi) = \nabla_\psi L(\phi,\psi) + \mu\psi \qquad \text{(IV.6.84)}$$

By theorem (IV.2.2), the bounds on the stationary value $L(\phi_e,\psi_e)$ are

$$\begin{aligned}
&\tfrac{1}{2}\delta<\phi_2,\phi_2> - \tfrac{1}{2}\mu<\psi_2,\psi_2> + L(\phi_4,\psi_2) - \tfrac{1}{2}\delta<\phi_4,\phi_4> \\
&\qquad + \tfrac{1}{2}\mu<\psi_2,\psi_2> - <\phi_2 - \phi_4, \delta\phi_2> \\
&\qquad \leq L(\phi_e,\psi_e) \leq \\
&\tfrac{1}{2}\delta<\phi_1,\phi_1> - \tfrac{1}{2}\mu<\psi_1,\psi_1> + L(\phi_1,\psi_3) - \tfrac{1}{2}\delta<\phi_1,\phi_1> \\
&\qquad + \tfrac{1}{2}\mu<\psi_3,\psi_3> - <\psi_1 - \psi_3, -\mu\psi_1>
\end{aligned} \qquad \text{(IV.6.85)}$$

where $\delta\phi_2 - \delta\phi_4 + \nabla_\phi L(\phi_4,\psi_2) = 0$ (IV.6.86)

and $-\mu\psi_1 + \mu\psi_3 + \nabla_\psi L(\phi_1,\psi_3) = 0$ (IV.6.87)

Using the last two equations, we have the bounds

$$\begin{aligned}
&-\tfrac{1}{2\delta}\,\|\nabla_\phi L(\phi_4,\psi_2)\|^2 + L(\phi_4,\psi_2) \\
&\qquad \leq L(\phi_e,\psi_e) \leq \\
&\tfrac{1}{2\mu}\,\|\nabla_\psi L(\phi_1,\psi_3)\|^2 + L(\phi_1,\psi_3)
\end{aligned} \qquad \text{(IV.6.88)}$$

To set up an iterative scheme, we let

$$\phi_1 = \phi_4 = \phi_{n-1}, \phi_2 = \phi_n, \quad \psi_2 = \psi_3 = \psi_{n-1}, \text{ and } \psi_1 = \psi_n \qquad \text{(IV.6.89)}$$

this results in the scheme

$$\phi_n = \phi_{n-1} - \frac{1}{\delta} \nabla_\phi L(\phi_{n-1}, \psi_{n-1}) \qquad \text{(IV.6.90)}$$

$$\psi_n = \psi_{n-1} + \frac{1}{\mu} \nabla_\psi L(\phi_{n-1}, \psi_{n-1}) \qquad \text{(IV.6.91)}$$

with the bounds

$$-\frac{1}{2\delta} \|\nabla_\phi L(\phi_{n-1}, \psi_{n-1})\|^2 \le L(\phi_e, \psi_e) - L(\phi_{n-1}, \psi_{n-1}) \le \frac{1}{2\mu} \|\nabla_\psi L(\phi_{n-1}, \psi_{n-1})\|^2 \qquad \text{(IV.6.92)}$$

It has not proved possible to show that the sequences $\{\phi_n\}$ and $\{\psi_n\}$ converge to the stationary value (ϕ_e, ψ_e); however, by proving that $\lim_{n\to\infty} \|\nabla_\phi L(\phi_n, \psi_n)\| = 0$ and $\lim_{n\to\infty} \|\nabla_\psi L(\phi_n, \psi_n)\| = 0$ and using the bounds given by equation (IV.6.92) we can prove the weaker result $\lim_{n\to\infty} L(\phi_n, \psi_n) = L(\phi_e, \psi_e)$ if certain conditions are placed on the gradients of $N(\phi, \psi)$; this is given in theorem (IV.6.1).

<u>Theorem (IV.6.1)</u>

Let $L(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ be a differentiable convex/concave saddle functional with a stationary value at (ϕ_e, ψ_e). Let $M(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ and $N(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ be convex/concave saddle functionals such that $L(\phi, \psi) = M(\phi, \psi) + N(\phi, \psi)$. Assume that sequences $\{\phi_n\}$ and $\{\psi_n\}$ can be constructed from equations (IV.6.90) and (IV.6.91).

Then, if there exists a real number γ, $0 < \gamma < \frac{1}{2}$, such that, for any pairs (ϕ_1, ψ_1) and (ϕ_2, ψ_2),

$$\|\nabla_\phi N(\phi_1, \psi_1) - \nabla_\phi N(\phi_2, \psi_2)\| \le \gamma\delta\|\phi_1 - \phi_2\| + \gamma\mu\|\psi_1 - \psi_2\| \qquad \text{(IV.6.93)}$$

and

$$\|\nabla_\psi N(\phi_1, \psi_1) - \nabla_\psi N(\phi_2, \psi_2)\| \le \gamma\delta\|\phi_1 - \phi_2\| + \gamma\mu\|\psi_1 - \psi_2\| \qquad \text{(IV.6.94)}$$

then $\lim_{n \to \infty} \|\nabla_\phi L(\phi_n, \psi_n)\| = 0$ (IV.6.95)

and

$\lim_{n \to \infty} \|\nabla_\psi L(\phi_n, \psi_n)\| = 0$ (IV.6.96)

<u>Proof</u>

$$\|\nabla_\phi L(\phi_n, \psi_n)\| = \|\nabla_\phi L(\phi_n, \psi_n) - \delta\phi_n + \delta\phi_n\|$$

$$= \|\nabla_\phi L(\phi_n, \psi_n) - \delta\phi_n + \delta\phi_{n-1} - \nabla_\phi L(\phi_{n-1}, \psi_{n-1})\|$$

(from equation (IV.6.90))

$$= \|\nabla_\phi N(\phi_n, \psi_n) - \nabla_\phi N(\phi_{n-1}, \psi_{n-1})\|$$

(from equation (IV.6.83))

$$\leq \gamma\delta\|\phi_n - \phi_{n-1}\| + \gamma\mu\|\psi_n - \psi_{n-1}\|$$

(from equation (IV.6.93))

$$= \gamma\delta\|\tfrac{1}{\delta}\nabla_\phi L(\phi_{n-1}, \psi_{n-1})\| + \gamma\mu\|\tfrac{1}{\mu}\nabla_\psi L(\phi_{n-1}, \psi_{n-1})\|$$

(from equations (IV.6.90) and (IV.6.91)).

That is, $\|\nabla_\phi L(\phi_n, \psi_n)\| \leq \gamma\|\nabla_\phi L(\phi_{n-1}, \psi_{n-1})\| + \gamma\|\nabla_\psi L(\phi_{n-1}, \psi_{n-1})\|$

(IV.6.97)

Similarly,

$$\|\nabla_\psi L(\phi_n, \psi_n)\| = \|\nabla_\psi L(\phi_n, \psi_n) + \mu\psi_n - \mu\psi_n\|$$

$$= \|\nabla_\psi L(\phi_n, \psi_n) + \mu\psi_n - \mu\psi_{n-1} - \nabla_\psi L(\phi_{n-1}, \psi_{n-1})\|$$

(from equation (IV.6.91))

$$= \|\nabla_\psi N(\phi_n, \psi_n) - \nabla_\psi N(\phi_{n-1}, \psi_{n-1})\|$$

(from equation (IV.6.84))

$$\leq \gamma\delta\|\phi_n - \phi_{n-1}\| + \gamma\mu\|\psi_n - \psi_{n-1}\|$$

(from equation (IV.6.94))

$$= \gamma\delta\|\tfrac{1}{\delta}\nabla_\phi L(\phi_{n-1}, \psi_{n-1})\| + \gamma\mu\|\tfrac{1}{\mu}\nabla_\psi L(\phi_{n-1}, \psi_{n-1})\|$$

(from equations (IV.6.90) and (IV.6.91))

That is, $\|\nabla_\psi L(\phi_n, \psi_n)\| \leq \gamma\|\nabla_\phi L(\phi_{n-1}, \psi_{n-1})\| + \gamma\|\nabla_\psi L(\phi_{n-1}, \psi_{n-1})\|$

(IV.6.98)

Therefore, putting equations (IV.6.97) and (IV.6.98) together gives

$$\|\nabla_\phi L(\phi_n, \psi_n)\| + \|\nabla_\psi L(\phi_n, \psi_n)\| \leq 2\gamma \left\{ \|\nabla_\phi L(\phi_{n-1}, \psi_{n-1})\| + \|\nabla_\psi L(\phi_{n-1}, \psi_{n-1})\| \right\}$$

and hence

$$\lim_{n\to\infty} \left\{ \|\nabla_\phi L(\phi_n, \psi_n)\| + \|\nabla_\psi L(\phi_n, \psi_n)\| \right\} = 0 \text{ as } 0 < \gamma < \tfrac{1}{2}$$

or $\lim_{n\to\infty} \|\nabla_\phi L(\phi_n, \psi_n)\| = 0$

and $\lim_{n\to\infty} \|\nabla_\psi L(\phi_n, \psi_n)\| = 0$

Substituting equations (IV.6.95) and (IV.6.96) into equation (IV.6.92) gives

$$0 \leq L(\phi_e, \psi_e) - \lim_{n\to\infty} L(\phi_{n-1}, \psi_{n-1}) \leq 0;$$

and hence if the functional $N(\phi, \psi)$ satisfies equations (IV.6.93) and (IV.6.94), then

$$\lim_{n\to\infty} L(\phi_n, \psi_n) = L(\phi_e, \psi_e) \qquad \text{(IV.6.99)}$$

In any application of the iterative scheme specified by equations (IV.6.90) and (IV.6.91), whether or not we can prove convergence of the scheme to $L(\phi_e, \psi_e)$ obviously depends on whether we can find a real number γ which satisfies equations (IV.6.93) and (IV.6.94); we now provide a simple example to show that these conditions can be satisfied.

Let $L(\phi, \psi) : E^1 \times E^1 \to \mathbb{R}$ be defined by the equation used in example 1 of this section,

$$L(\phi, \psi) = \tfrac{1}{2}\langle \phi, B\phi \rangle - \tfrac{1}{2}\langle \psi, B\psi \rangle + \langle f, \phi \rangle + \langle g, \psi \rangle \qquad \text{(IV.6.100)}$$

where B is a linear, symmetric, positive definite operator. Then from equation (IV.6.80),

$$N(\phi, \psi) = \tfrac{1}{2}\langle \phi, (B - \delta I)\phi \rangle - \tfrac{1}{2}\langle \psi, (B - \mu I)\psi \rangle + \langle \phi, f \rangle + \langle \psi, g \rangle \qquad \text{(IV.6.101)}$$

where we require $B - \delta I \geqslant 0$ and $B - \mu I \geqslant 0$ (IV.6.102)

to ensure that $N(\phi, \psi)$ is a convex/concave saddle functional. Without any loss of generality we can assume that $\mu = \delta$. Then,

$$\|\nabla_\phi N(\phi_1, \psi_1) - \nabla_\phi N(\phi_2, \psi_2)\|$$
$$= \|(B - \delta I)(\phi_1 - \phi_2)\|$$
$$\leq \gamma\delta \|\phi_1 - \phi_2\| + \gamma\delta \|\psi_1 - \psi_2\|$$

if $(B - \delta I) \leq \gamma\delta I$ (IV.6.103)

Similarly,

$$\|\nabla_\psi N(\phi_1, \psi_1) - \nabla_\psi N(\phi_2, \psi_2)\|$$
$$= \|(B - \delta I)(\psi_1 - \psi_2)\|$$
$$\leq \gamma\delta \|\phi_1 - \phi_2\| + \gamma\delta \|\psi_1 - \psi_2\|$$

if again $(B - \delta I) \leq \gamma\delta I$.

Let $\delta = 1$ and $B = I + K$, where K is the integral operator which was considered in example 1;

$$K\phi(x) = \int_0^1 k(x,y)\phi(y)\, dy$$

where $k(x,y) = \begin{cases} x(1-y), & x \leq y \\ y(1-x), & x > y \end{cases}$ (IV.6.104)

From example 1 of section II.15, $0 < K \leq \frac{1}{\pi^2} I$ (IV.6.105)

$$B - \delta I - \gamma\delta I = K + I - I - \gamma I$$
$$= K - \gamma I$$
$$\leq 0 \text{ if } K \leq \gamma I$$

We can take $\gamma = \frac{1}{\pi^2}$, as $K < \frac{1}{\pi^2} I$; as $\gamma = \frac{1}{\pi^2} < \frac{1}{2}$,

then by theorem (IV.6.1), the iterative scheme defined by the equations

$$\phi_n = - K\phi_{n-1} - f, \quad \text{(IV.6.106)}$$

$$\psi_n = - K\psi_{n-1} + g \quad \text{(IV.6.107)}$$

$$L(\phi_n, \psi_n) = \int_0^1 \left\{ \tfrac{1}{2}\phi_n (K + I)\phi_n - \tfrac{1}{2}\psi_n (K + I)\psi_n + \phi_n f + \psi_n g \right\} dx \quad \text{(IV.6.108)}$$

will converge to the stationary value

$$L(\phi_e, \psi_e) = \tfrac{1}{2} \int_0^1 \{\phi_e f + \psi_e g\}\, dx \quad \text{(IV.6.109)}$$

where $(K + I)\phi_e + f = 0$ and $-(K + I)\psi_e + g = 0$ (IV.6.110)

CHAPTER IV

Example 4

We are going to apply theorem (IV.2.2) to the boundary value problem given by the equation

$$\nabla^2 u + F'(u) = 0 \text{ in } V \quad u = u_s \text{ on } S \tag{IV.6.111}$$

where V is a region in $\mathbb{R}^2$ bounded by a piecewise smooth curve S, u_s is specified on S and F(u) is a given concave differentiable function.

To accommodate the boundary term $u = u_s$, we let

$$\phi = \begin{Bmatrix} u \text{ in } V \\ u \text{ on } S \end{Bmatrix} \quad \text{and} \quad \psi = \begin{Bmatrix} \underline{v} \text{ in } V \\ \underline{v} \text{ on } S \end{Bmatrix} \tag{IV.6.112}$$

where $\underline{v}$ is a vector field. The inner products are defined by

$$\langle \phi_1, \phi_2 \rangle = \int_V u_1\, u_2\, dt + \int_S u_1\, u_2\, dt \tag{IV.6.113}$$

and

$$\langle \psi_1, \psi_2 \rangle = \int_V \underline{v}_1\, \underline{v}_2\, dt + \int_S \underline{v}_1\, \underline{v}_2\, dt \tag{IV.6.114}$$

where line integral boundary inner product terms are included.

To achieve the flexibility required to apply theorem (IV.2.2) we will define the convex/concave saddle functional $L(\phi, \psi)$ by

$$L(\phi, \psi) = \int_V \Big\{ \tfrac{1}{2}(1 - q^2 r^{-1})(\text{grad } u)^2 - q\, u \text{ div } \underline{v} - \tfrac{1}{2} r\, \underline{v}.\underline{v} - F(u) \Big\}\, d\tau + \int_S \Big\{ \tfrac{1}{2} p_s (u - u_s)^2 - \big((rq^{-1} - q)\, u - rq^{-1} u_s\big)\, \underline{n}.\underline{v} \Big\}\, d\sigma \tag{IV.6.115}$$

where $\underline{n}$ is the outward unit normal to S, q and r are real non-zero numbers and p_s is a continuous function defined on S. The functional derivatives of $L(\phi, \psi)$ are then

$$\nabla_\phi L(\phi, \psi) = \begin{bmatrix} -(1 - q^2 r^{-1})\nabla^2 u - q \text{ div } \underline{v} - F'(u) \text{ in } V \\ p_s (u - u_s) - (rq^{-1} - q)\, \underline{n}.(\underline{v} - qr^{-1} \text{ grad } u) \text{ on } S \end{bmatrix} \tag{IV.6.116}$$

and
$$\nabla_\psi L(\phi, \psi) = \begin{bmatrix} q \text{ grad } u - r\underline{v} \text{ in } V \\ -rq^{-1}(u - u_s)\, \underline{n} \text{ on } S \end{bmatrix} \tag{IV.6.117}$$

It can easily be shown, using equation (II.9.1), that $L(\phi, \psi)$ is a strict convex/concave saddle functional if

$$(1 - q^2 r^{-1}) \geqslant 0, \quad p_s \geqslant 0 \quad \text{and } r > 0 \tag{IV.6.118}$$

The stationary value of $L(\phi_e, \psi_e)$ occurs at $(u_e, \underline{v}_e)$, when

$$\nabla_\phi L(\phi, \psi) = \nabla_\psi L(\phi, \psi) = 0; \text{ that is,} \tag{IV.6.119}$$

$$(1 - q^2 r^{-1}) \nabla^2 u_e + q \operatorname{div} \underline{v}_e + F'(u_e) = 0 \text{ in } V \tag{IV.6.120}$$

$$p_s (u_e - u_s) - (rq^{-1} - q)\, \underline{n} \cdot (\underline{v}_e - qr^{-1} \operatorname{grad} u_e) = 0 \text{ in } S \tag{IV.6.121}$$

$$q \operatorname{grad} u_e - r\, \underline{v}_e = 0 \text{ in } V \tag{IV.6.122}$$

$$rq^{-1} (u_e - u_s) = 0 \text{ on } S$$

Eliminating v_e from the above set of equations results in equation (IV.6.111), as required. The classical dual extremum principles are

$$L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha) \tag{IV.6.123}$$

where
$$L(\phi_\beta, \psi_\beta) = \int_V \Big\{ -\tfrac{1}{2} (1 - q^2 r^{-1}) (\operatorname{grad} u_\beta)^2 + u_\beta F'(u_\beta) - \tfrac{1}{2} r\, \underline{v}_\beta \cdot \underline{v}_\beta - F(u_\beta) \Big\} d\tau + \int_S \Big\{ -\tfrac{1}{2} p_s (u_\beta^2 - u_s^2) + u_s rq^{-1} \underline{n} \cdot \underline{v}_\beta \Big\} d\sigma \tag{IV.6.124}$$

$$\text{with} \begin{bmatrix} -(1 - q^2 r^{-1}) \nabla^2 u_\beta - q \operatorname{div} \underline{v}_\beta - F'(u_\beta) = 0 \text{ in } V \\ p_s (u_\beta - u_s) - (rq^{-1} - q)\, \underline{n} \cdot (\underline{v}_\beta - qr^{-1} \operatorname{grad} u_\beta) = 0 \text{ in } S \end{bmatrix} \tag{IV.6.125}$$

$$\begin{aligned} L(\phi_\alpha, \psi_\alpha) &= \int_V \Big\{ \tfrac{1}{2} (1 - q^2 r^{-1}) (\operatorname{grad} u_\alpha)^2 + \tfrac{1}{2} r\, \underline{v}_\alpha \cdot \underline{v}_\alpha - F(u_\alpha) \Big\} d\tau \\ &= \int_V \Big\{ \tfrac{1}{2} (\operatorname{grad} u_\alpha)^2 - F(u_\alpha) \Big\} d\tau \end{aligned} \tag{IV.6.126}$$

$$\text{with} \begin{bmatrix} q \operatorname{grad} u_\alpha - r\underline{v}_\alpha = 0 \text{ in } V \\ -rq^{-1} (u_\alpha - u_s)\, \underline{n} = 0 \text{ in } S \end{bmatrix} \tag{IV.6.127}$$

and

$$L(\phi_e, \psi_e) = \int_V \Big\{ -F(u_e) + \tfrac{1}{2} u_e F'(u_e) \Big\} d\tau + \int_S \tfrac{1}{2} u_s\, \underline{n} \cdot \operatorname{grad} u_e \, d\sigma \tag{IV.6.128}$$

The stationary value $L(\phi_e, \psi_e)$ and upper bound $L(\phi_\alpha, \psi_\alpha)$ do not include any of the parameters q, r and p_s and hence these are the same as the usual stationary value and upper bound. If $1 - q^2 r^{-1} = 0$, $r = 1$ and $p_s = 0$, the lower bound reduces to the usual lower bound; to avoid this we will change equation (IV.6.118) to:

we require $1 - q^2 r^{-1} > 0$ and $p_s > 0$ (IV.6.129)

We are now going to find a decomposition lower bound using theorem (IV.2.2). By this theorem we have to find convex/concave saddle functionals $M(\phi, \psi)$ and $N(\phi, \psi)$, with at least one strictly convex and at least one strictly concave, such that $L(\phi, \psi) = M(\phi, \psi) + N(\phi, \psi)$.

The decomposition lower bound L_D is given by

$$L_D = M(\phi_2, \psi_2) + N(\phi_4, \psi_2) - \langle \phi_2 - \phi_4, \ \nabla_\phi M(\phi_2, \psi_2) \rangle \qquad \text{(IV.6.130)}$$

where $\nabla_\phi M(\phi_2, \psi_2) + \nabla_\phi N(\phi_4, \psi_2) = 0$ (IV.6.131)

$$\text{Let } M(\phi, \psi) = \tfrac{1}{2}\int_V \lambda_V u^2 \, d\tau + \tfrac{1}{2}\int_S \lambda_s u^2 \, d\sigma \qquad \text{(IV.6.132)}$$

where λ_V and λ_s are real numbers greater than zero; then $M(\phi, \psi)$ is a strict convex/concave saddle functional. We then require $N(\phi, \psi)$ to be equal to $L(\phi, \psi) - M(\phi, \psi)$; that is

$$N(\phi, \psi) = \int_V \left\{ -\tfrac{1}{2} u \left(\lambda_V u + (1 - q^2 r^{-1}) \nabla^2 u \right) - q\, u \operatorname{div} \underline{v} - \tfrac{1}{2} r\, \underline{v} \cdot \underline{v} - F(u) \right\} d\tau$$
$$+ \int_S \left\{ \tfrac{1}{2} u \left(-\lambda_s u + (1 - q^2 r^{-1})\, \underline{n} \cdot \nabla u \right) + \tfrac{1}{2} p_s (u - \right.$$

(IV.6.133)

$N(\phi, \psi)$ will be a weak convex/concave saddle functional, which is all that we require, if, using equation (II.9.1),

$$\tfrac{1}{2}\int_V \left\{ (1 - q^2 r^{-1}) (\operatorname{grad}(u_1 - u_2))^2 - \lambda_V (u_1 - u_2)^2 + r\, (\underline{v}_1 - \underline{v}_2)^2 \right\} d\tau$$
$$+ \tfrac{1}{2}\int_S (p_s - \lambda_s)(u_1 - u_2)^2 \, d\sigma \geq 0 \qquad \text{(IV.6.134)}$$

Essentially, equation (IV.6.134) requires that we construct a lower bound for $\int_V (\nabla(u_1 - u_2))^2 \, d\tau$; this can be supplied by Friedrich's inequality, from (44). We construct a rectangle $a_1 \leq x \leq a_2$, $b_1 \leq y \leq b_2$, such that all points of S are inside the rectangle.

Then, for any function u with partial derivatives on S,

$$\pi^2 K \int_V u^2 \, d\tau \leq \int_V (\text{grad } u)^2 \, d\tau + C \int_S u^2 \, d\sigma \qquad \text{(IV.6.135)}$$

$$\text{where } K = \frac{1}{(a_2 - a_1)^2} + \frac{1}{(b_2 - b_1)^2} \qquad \text{(IV.6.136)}$$

$$\text{and } C = \sup_{(x,y) \in S} \frac{|n \cdot \text{grad } f|}{f} \qquad \text{(IV.6.137)}$$

$$\text{where } f = \sin \frac{\pi (x - a_1)}{(a_2 - a_1)} \sin \frac{\pi (y - b_1)}{(b_2 - b_1)} \qquad \text{(IV.6.138)}$$

The strict enclosure of S ensures that f does not vanish in D or on S, and S can be piecewise smooth if points of discontinuous slope are deleted in the definition of C in equation (IV.6.137).

Using equations (IV.6.134) and (IV.6.135),

$$\frac{1}{2} \int_V \left\{ (1 - q^2 r^{-1}) (\nabla(u_1 - u_2))^2 - \lambda_V (u_1 - u_2)^2 + r (v_1 - v_2)^2 \right\} d\tau$$

$$+ \frac{1}{2} \int_S (p_s - \lambda_s) (u_1 - u_2)^2 \, d\sigma$$

$$\geq \frac{1}{2} \int_V \left\{ \pi^2 K(1 - q^2 r^{-1}) - \lambda_V (u_1 - u_2)^2 + r (v_1 - v_2)^2 \right\} d\tau$$

$$+ \frac{1}{2} \int_S \left(p_s - \lambda_s - C (1 - q^2 r^{-1}) \right) (u_1 - u_2)^2 \, d\sigma$$

$$\geq 0 \text{ if } \pi^2 K(1 - q^2 r^{-1}) - \lambda_V \geq 0 \qquad \text{(IV.6.139)}$$

$$\text{and } p_s - \lambda_s - C(1 - q^2 r^{-1}) \geq 0$$

We therefore let $\lambda_V = \pi^2 K(1 - q^2 r^{-1})$ (IV.6.140)

$$\text{and } \lambda_s = \inf_{(x,y) \in S} p_s - C(1 - q^2 r^{-1}) \qquad \text{(IV.6.141)}$$

where the right side of equation (IV.6.141) is positive.

With this choice of $M(\phi,\psi)$ and $N(\phi,\psi)$, the gradient equation (IV.6.140) becomes

$$\left[\begin{array}{l} \lambda_v (u_2 - u_4) - (1 - q^2 r^{-1})\nabla^2 u_4 - q \operatorname{div} \underline{v}_2 - F'(u_4) = 0 \text{ in } V \\ \lambda_s (u_2 - u_4) + p_s (u_4 - u_s) - (rq^{-1} - q)\, \underline{n}.\, (\underline{v}_2 - qr^{-1}\nabla u_4) = 0 \text{ on } S \end{array}\right] \tag{IV.6.142}$$

and the decomposition lower bound is, using the last equation

$$\begin{aligned} L_D = \int_V \Big\{ &- \tfrac{1}{2} \lambda_v (u_2^2 - u_4^2) - \tfrac{1}{2} (1 - q^2 r^{-1}) (\operatorname{grad} u_4)^2 \\ &- \tfrac{1}{2} r\underline{v}_2 . \underline{v}_2 - F(u_4) + u_4 F'(u_4) \Big\} d\tau \\ + \int_S \Big\{ &- \tfrac{1}{2} \lambda_s (u_2^2 - u_4^2) - \tfrac{1}{2} p_s (u_4^2 - u_s^2) \\ &+ rq^{-1} u_s \underline{n}. \underline{v}_2 \Big\} d\sigma \end{aligned} \tag{IV.6.143}$$

It is easy to see that if $u_2 = u_4$, L_D reduces to the lower bound given in equation (IV.6.124).

We are going to illustrate this application of the decomposition dual extremum principles by using the classical torsion problem in which

$$F(u) = u \quad \text{and } u_s = 0 \tag{IV.6.144}$$

After finding decomposition lower bounds for two particular regions, these will be compared with the relevant stationary values and also with bounds obtained using a different method.

<u>Decomposition Lower Bound</u>

From equations (IV.6.142), (IV.6.143) and (IV.6.144),

$$\begin{aligned} L_D = \int_V \Big\{ &- \tfrac{1}{2} \lambda_v (u_2^2 - u_4^2) - \tfrac{1}{2} (1 - q^2 r^{-1}) (\operatorname{grad} u_4)^2 \\ &- \tfrac{1}{2} r\, \underline{v}_2 . \underline{v}_2 \Big\} d\tau \\ + \int_S \Big\{ &- \tfrac{1}{2} \lambda_s (u_2^2 - u_4^2) - \tfrac{1}{2} p_s u_4^2 \Big\} d\sigma \end{aligned} \tag{IV.6.145}$$

where $\lambda_v (u_2 - u_4) - (1 - q^2 r^{-1})\nabla^2 u_4 - q \operatorname{div} \underline{v}_2 - 1 = 0$ in V (IV.6.146)

and $\lambda_s (u_2 - u_4) + p_s u_4 - (rq^{-1} - q)\, \underline{n}.\, (\underline{v}_2 - qr^{-1}\nabla u_4) = 0$ on S

(IV.6.147)

In order to simplify the bound, we are going to let

$$\underline{v}_2 = q r^{-1} \text{ grad } u_4 \quad \text{(IV.6.148)}$$

Equations (IV.6.146) and (IV.6.147) then become

$$u_2 = u_4 + \lambda_v^{-1} (\nabla^2 u_4 + 1) \text{ in } V \quad \text{(IV.6.149)}$$

and $u_2 = u_4 - \lambda_s^{-1} p_s u_4$ on S (IV.6.150)

with
$$L_D = \int_V \left\{ - \tfrac{1}{2} \lambda_v (u_2^2 - u_4^2) - \tfrac{1}{2} (\text{grad } u_4)^2 \right\} d\tau + \int_S \left\{ - \tfrac{1}{2} \lambda_s (u_2^2 - u_4^2) - \tfrac{1}{2} p_s u_4^2 \right\} d\sigma$$
$$= -\int_V \left\{ \tfrac{1}{2} \lambda_v^{-1} (\nabla^2 u_4 + 1)^2 + \tfrac{1}{2} (\nabla u_4)^2 + u_4 \nabla^2 u_4 + u_4 \right\} d\tau - \int_S \left\{ \tfrac{1}{2} \lambda_s^{-1} p_s^2 - p_s \right\} u_4^2 \, d\sigma \quad \text{(IV.6.151)}$$

where u_2 has been eliminated using equations (IV.6.149) and (IV.6.150). We cannot choose u_4 completely without restriction as the boundary value of u_2 derived from equation (IV.6.149) must be compatible with equation (IV.6.150) that is u_4 must satisfy

$$\lambda_v^{-1} (\nabla^2 u_4 + 1) = - \lambda_s^{-1} p_s u_4 \text{ on } S \quad \text{(IV.6.152)}$$

For a trial function we will choose $u_4 = B$, (IV.6.153)

a constant and, from equation (IV.6.141), let

$$\lambda_s = p_s - C(1 - \gamma) \quad \text{(IV.6.154)}$$

where p_s is a constant and $\gamma = q^2 r^{-1}$ (IV.6.155)

As $\lambda_v = \pi^2 K(1 - \gamma)$, from equation (IV.6.140), B becomes

$$B = - \frac{(p_s - C(1 - \gamma))}{p_s \pi^2 K(1 - \gamma)} \quad \text{(IV.6.156)}$$

Substituting this equation for B, and those for λ_s and λ_v into equation (IV.6.151) results in

$$L_D = \int_V \left\{ \frac{1}{2 \pi^2 K(1 - \gamma)} - \frac{C}{p_s \pi^2 K} \right\} d\tau - \int_S \left\{ \frac{C}{2 \pi^4 K^2 (1 - \gamma)} - \frac{C^2}{p_s \pi^4 K^2} \right\} d\sigma \quad \text{(IV.6.157)}$$

If A is the area enclosed by S and l its length, integration of the above equation gives

$$L_D = \frac{1}{2\pi^4 K^2}\left((2AC\pi^2 K - \ell C^2)\frac{1}{p_s} + (\ell C - A\pi^2 K)\frac{1}{1-\gamma}\right) \quad \text{(IV.6.158)}$$

To find the best decomposition lower bound, this expression for L_D has to be maximised subject to $0 < \gamma < 1$ and $p_s > C(1-\gamma)$ (IV.6.159)

This is a linear programming problem in terms of the reciprocals of $(1-\gamma)$ and p_s. As can be seen from figure (IV.6.1), the feasible points occur at $\gamma = 0$, $p_s = C$ and $\gamma = 0$, $p_s^{-1} \to 0$.

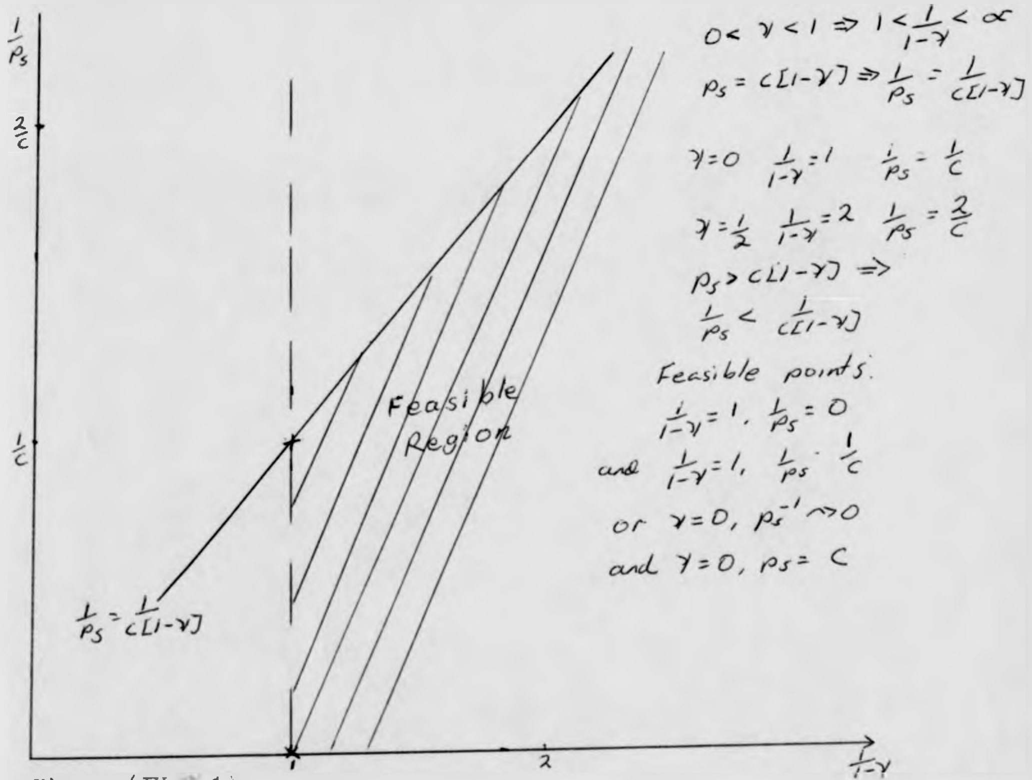


Figure (IV.6.1)

Neither of these feasible points is strictly attainable, but they can be approached as closely as we wish.

At $\gamma = 0$, $p_s = C$, $L_D = -\dfrac{A}{2\pi^2 K}$ (IV.6.160)

and at $\gamma = 0$, $p_s^{-1} \to 0$, $L_D = -\dfrac{(\ell C - A\pi^2 K)}{2\pi^4 K^2}$ (IV.6.161)

Although it cannot be proved that the optimum upper bound is given by equation (IV.6.160), it is logical to assume that this is so, as the bound in equation

(IV.6.161) assumes an unbounded value for p_s.

We therefore have $L(\phi_e, \psi_e) \geqslant - \dfrac{A}{2\pi^2 K}$ (IV.6.162)

This inequality holds for all rectangles which enclose V. The maximum value of the lower bound must occur when $K = \dfrac{1}{d_1^2} + \dfrac{1}{d_2^2}$ where d_1 and d_2 are the lengths of the sides of the smallest rectangle enclosing D; this is illustrated in figure (IV.6.2).

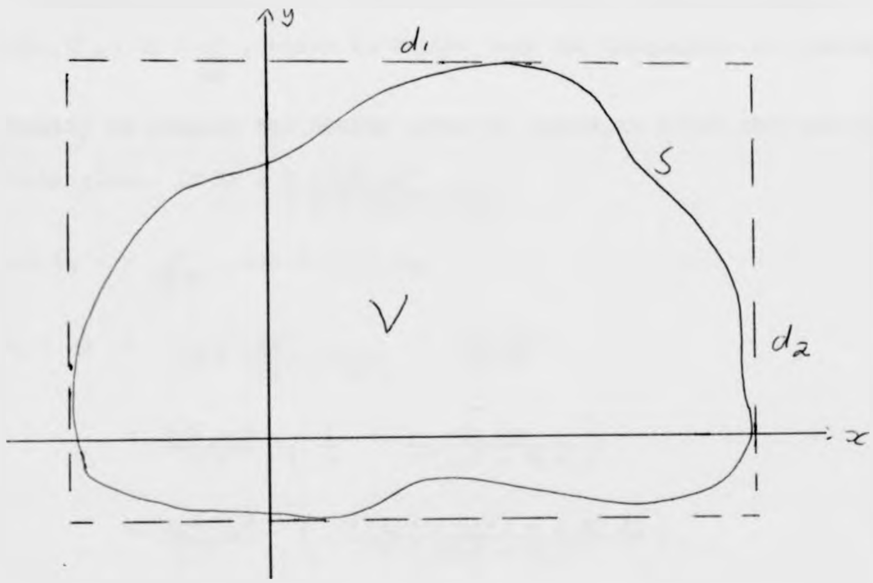


Figure (IV.6.2)

Substituting this value for K into equation (IV.6.52) gives

$$L(\phi_e, \psi_e) \geqslant - \frac{A d_1^2 d_2^2}{2\pi^2 (d_1^2 + d_2^2)} \quad \text{(IV.6.163)}$$

If V is a circle of radius c, we have $A = \pi c^2$ and $d_1 = d_2 = 2c$.

Hence $L(\phi_e, \psi_e) \geqslant - \dfrac{c^4}{\pi}$ (IV.6.164)

If V is a square $-c < x < c$, $-c < y < c$ then $A = 4c^2$ and $d_1 = d_2 = 2c$; hence $L(\phi_e, \psi_e) \geqslant - \dfrac{4c^4}{\pi^2}$ (IV.6.165)

The stationary values are, for the circle, $L(\phi_e, \psi_e) = -\dfrac{\pi c^4}{16}$ and for the square, $L(\phi_e, \psi_e) = -0.281\ldots c^4$ (see page 278 of (73)).

From page 194 of (70), the Saint Venant inequality for this problem is

$$L(\phi_e, \psi_e) \geqslant -\frac{A^2}{16\pi} \qquad \text{(IV.6.166)}$$

For the circle, this gives $L(\phi_e, \psi_e) \geqslant -\dfrac{\pi c^4}{16}$, which is the actual stationary value, and for the square the Saint Venant inequality is

$L(\phi_e, \psi_e) \geqslant -\dfrac{c^4}{16}$, which is better than the inequality in equation (IV.6.165).

Finally we compare the bounds given by equations (IV.6.163) and (IV.6.166) for rectangles. If $L_1 = \dfrac{-Ad_1^2\, d_2^2}{2\pi^2 (d_1^2 + d_2^2)}$

and $L_2 = -\dfrac{A^2}{16\pi}$, and $A = d_1\, d_2$,

$$L_1 - L_2 = \frac{-d_1^3\, d_2^3}{2\pi^2 (d_1^2 + d_2^2)} + \frac{d_1^2\, d_2^2}{16\pi}$$

$$= \frac{d_1^2\, d_2^2}{2\pi}\left(\frac{1}{8} - \frac{d_1\, d_2}{\pi(d_1^2 + d_2^2)}\right)$$

$$= \frac{d_1^2\, d_2^2}{2\pi}\left(\frac{\pi(d_1^2 + d_2^2) - 8\, d_1\, d_2}{8\pi(d_1^2 + d_2^2)}\right)$$

$$= \frac{d_1^2\, d_2^2}{16\pi(d_1^2 + d_2^2)}\left(d_1 - \left(\frac{4}{\pi} + \frac{\sqrt{16 - \pi^2}}{\pi}\right) d_2\right)\left(d_1 - \left(\frac{4}{\pi} - \frac{\sqrt{16 - \pi^2}}{\pi}\right) d_2\right)$$

$$> 0 \text{ if } \frac{d_1}{d_2} > \left(\frac{4}{\pi} + \frac{\sqrt{16 - \pi^2}}{\pi}\right) \text{ or } \frac{d_1}{d_2} < \left(\frac{4}{\pi} - \frac{\sqrt{16 - \pi^2}}{\pi}\right)$$

That is, the comparison bounds given in equation (IV.6.163) is better than the Saint Venant bound given in equation (IV.6.166) if the ratio of the sides is less than about 0.485 or greater than about 2.061.

CHAPTER V

V.1 Introduction

This chapter is on comparison functionals. These were briefly discussed in section I.7 of the survey, where two papers which form the basis for the ideas in this chapter were reviewed. This chapter expands on and extends the work carried out in (68). Part of this chapter, in section V.2, was originally written in conjunction with P Smith for paper (15).

The basic idea is that if we are given a saddle functional $L(\phi, \psi)$, we try to find other functionals containing simpler operators than those in $L(\phi, \psi)$ which can be used to find upper and lower bounds to $L(\phi_e, \psi_e)$. It is possible that the inclusion of these simpler operators could enable bounds to $L(\phi_e, \psi_e)$ to be found more easily than those obtained using the usual dual extremum principles. Very simple comparison operators, such as multiples of the identity operator, can produce good results.

We start in section V.2 by setting out and proving the theorems involving comparison functionals. In section V.3, two of these theorems are applied to the usual quadratic functional and conditions on the comparison operators and trial functions in the comparison bounds are found which guarantee that the comparison bounds are sharper than the classical bounds.

Section V.4 considers iterative methods. Cobweb iterative methods are applied to the comparison dual extremum principles obtained in section V.3 by applying the two theorems to the quadratic functional. Conditions for convergence are found and simple examples are provided to show that these conditions can be met. The section ends with a brief discussion on iterative methods for problems for which we cannot prove convergence.

The last section, section V.5, deals with two applications of the methods developed in the previous sections.

CHAPTER V

V.2 Comparison Dual Extremum Principles

This section sets out the basic results on which this chapter is based. After some definitions, a lemma necessary for proofs of the theorems in this section is given, and then the theorems themselves. This section is an expanded version of the main part of the jointly written paper, (15).

In each of the theorems we will be finding bounds to the real convex/concave saddle functional $L(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$, where $E^1 \times F^1$ is a product space of two vector spaces. We shall assume that the partial functional derivatives $\nabla_\phi L(\phi, \psi)$ and $\nabla_\psi L(\phi, \psi)$ exist; definition (II.7.2) sets out the conditions for these derivatives to exist and also how they can be found.

Properties of functionals which we will need in this section are, denoting a general functional as $T(\phi, \psi)$, that $T(\phi, \psi)$ is convex or concave in ϕ or ψ , jointly convex in ϕ and ψ , and jointly concave in ϕ and ψ . These properties are defined using the functional derivatives in definitions (II.8.4) to (II.8.6); as they will be used extensively in the proofs in this section, for clarity they will be briefly summarised here.

Definition (V.2.1)

Assume that the functional $T(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ has a gradient with respect to ϕ at all points of $C \subseteq E^1$ and a gradient with respect to ψ at all points $D \subseteq F^1$, then

(i) $T(\phi, \psi)$ is convex with respect to ϕ if

$$T(\phi_1, \psi) - T(\phi_2, \psi) - \langle \phi_1 - \phi_2, \nabla_\phi T(\phi_2, \psi) \rangle \geq 0$$
$$\forall (\phi_1, \phi_2) \in C \text{ and } \forall \psi \in F^1. \qquad (V.2.1)$$

(ii) $T(\phi, \psi)$ is concave with respect to ϕ if

$$T(\phi_1, \psi) - T(\phi_2, \psi) - \langle \phi_1 - \phi_2, \nabla_\phi T(\phi_1, \psi) \rangle \geq 0$$
$$\forall (\phi_1, \phi_2) \in C \text{ and } \forall \psi \in F^1. \qquad (V.2.2)$$

(iii) $T(\phi,\psi)$ is convex with respect to ψ if

$$T(\phi,\psi_1) - T(\phi,\psi_2) - \langle \psi_1 - \psi_2, \nabla_\psi T(\phi,\psi_2)\rangle \geq 0$$
$$\forall \phi \in E^1 \text{ and } \forall (\psi_1,\psi_2) \in D \tag{V.2.3}$$

(iv) $T(\phi,\psi)$ is concave with respect to ψ if

$$T(\phi,\psi_1) - T(\phi,\psi_2) - \langle \psi_1 - \psi_2, \nabla_\psi T(\phi,\psi_1)\rangle \geq 0$$
$$\forall \phi \in E^1 \text{ and } \forall (\psi_1,\psi_2) \in D \tag{V.2.4}$$

(v) $T(\phi,\psi)$ is jointly convex with respect to ϕ and ψ if

$$- T(\phi_1,\psi_1) + T(\phi_2,\psi_2) + \langle \phi_1 - \phi_2, \nabla_\phi T(\phi_1,\psi_1)\rangle$$
$$+ \langle \psi_1 - \psi_2, \nabla_\psi T(\phi_1,\psi_1)\rangle \geq 0$$
$$\forall (\phi_1,\psi_1) \text{ and } (\phi_2,\psi_2) \in C \times D \tag{V.2.5}$$

and finally

(vi) $T(\phi,\psi)$ is jointly concave with respect to ϕ and ψ if

$$- T(\phi_1,\psi_1) + T(\phi_2,\psi_2) + \langle \phi_1 - \phi_2, \nabla_\phi T(\phi_2,\psi_2)\rangle$$
$$+ \langle \psi_1 - \psi_2, \nabla_\psi T(\phi_2,\psi_2)\rangle \geq 0$$
$$\forall (\phi_1,\psi_1) \text{ and } (\phi_2,\psi_2) \in C \times D \tag{V.2.6}$$

The basis of the theorems which follow is a result which compares two functionals; the general version of this comparison is given in the following lemma.

<u>Lemma (V.2.1)</u>

Let $P(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ be differentiable with respect to ϕ for all $(\phi,\psi) \in E^1 \times F^1$, and be convex in ϕ, and let $Q(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ be differentiable with respect to ψ for all $(\phi,\psi) \in E^1 \times F^1$, and be concave in ψ.

Let (ϕ_1,ψ_1) and (ϕ_2,ψ_2) be any pairs which satisfy

$$\nabla_\phi P(\phi_1,\psi_1) = 0 \quad \text{and} \quad \nabla_\psi Q(\phi_2,\psi_2) = 0 \tag{V.2.7}$$

Then, if $P(\phi,\psi) \leq Q(\phi,\psi) \quad \forall \; (\phi,\psi) \in E^1 \times F^1$ (V.2.8)

it follows that $P(\phi_1,\psi_1) \leq Q(\phi_2,\psi_2)$ (V.2.9)

Proof

Using equation (V.2.3), with ϕ and ψ interchanged and $\psi = \psi_1$, as $P(\phi, \psi)$ is convex in ϕ we have

$P(\phi_2, \psi_1) - P(\phi_1, \psi_1) - \langle \phi_2 - \phi_1, \nabla_\phi P(\phi_1, \psi_1) \rangle \geqslant 0$

or, as $\nabla_\phi P(\phi_1, \psi_1) = 0$,

$P(\phi_1, \psi_1) \leq P(\phi_2, \psi_1)$

Hence from equation (V.2.8),

$$P(\phi_1, \psi_1) \leq P(\phi_2, \psi_1) \leq Q(\phi_2, \psi_1) \qquad \text{(V.2.10)}$$

Using equation (V.2.4) with ψ_1 and ψ_2 interchanged and $\phi = \phi_2$, as $Q(\phi, \psi)$ is concave in ψ we have

$Q(\phi_2, \psi_2) - Q(\phi_2, \psi_1) - \langle \psi_2 - \psi_1, \nabla_\psi Q(\phi_2, \psi_2) \rangle \geqslant 0$

or as $\nabla_\psi Q(\phi_2, \psi_2) = 0$,

$$Q(\phi_2, \psi_1) \leq Q(\phi_2, \psi_2) \qquad \text{(V.2.11)}$$

Putting equations (V.2.10) and (V.2.11) together gives

$P(\phi_1, \psi_1) \leq Q(\phi_2, \psi_2)$ as required.

If, in addition $P(\phi, \psi)$ is concave in ψ and $Q(\phi, \psi)$ is convex in ϕ, then $P(\phi, \psi)$ and $Q(\phi, \psi)$ are both convex/concave saddle functionals; if (ϕ_1, ψ_1) and (ϕ_2, ψ_2) are stationary points of $P(\phi, \psi)$ and $Q(\phi, \psi)$ respectively, then it is obvious that in the lemma we are comparing stationary values. This version was proved by Smith in (68). Geometrically, this implies that if one saddle surface is above another then their stationary values must also satisfy this property. This is illustrated in figure (V.2.1).

Proof

Using equation (V.2.3), with ϕ and ψ interchanged and $\psi = \psi_1$, as $P(\phi, \psi)$ is convex in ϕ we have

$$P(\phi_2, \psi_1) - P(\phi_1, \psi_1) - \langle \phi_2 - \phi_1, \nabla_\phi P(\phi_1, \psi_1)\rangle \geqslant 0$$

or, as $\nabla_\phi P(\phi_1, \psi_1) = 0$,

$$P(\phi_1, \psi_1) \leq P(\phi_2, \psi_1)$$

Hence from equation (V.2.8),

$$P(\phi_1, \psi_1) \leq P(\phi_2, \psi_1) \leq Q(\phi_2, \psi_1) \qquad (V.2.10)$$

Using equation (V.2.4) with ψ_1 and ψ_2 interchanged and $\phi = \phi_2$, as $Q(\phi, \psi)$ is concave in ψ we have

$$Q(\phi_2, \psi_2) - Q(\phi_2, \psi_1) - \langle \psi_2 - \psi_1, \nabla_\psi Q(\phi_2, \psi_2)\rangle \geqslant 0$$

or as $\nabla_\psi Q(\phi_2, \psi_2) = 0$,

$$Q(\phi_2, \psi_1) \leq Q(\phi_2, \psi_2) \qquad (V.2.11)$$

Putting equations (V.2.10) and (V.2.11) together gives

$P(\phi_1, \psi_1) \leq Q(\phi_2, \psi_2)$ as required.

If, in addition $P(\phi, \psi)$ is concave in ψ and $Q(\phi, \psi)$ is convex in ϕ, then $P(\phi, \psi)$ and $Q(\phi, \psi)$ are both convex/concave saddle functionals; if (ϕ_1, ψ_1) and (ϕ_2, ψ_2) are stationary points of $P(\phi, \psi)$ and $Q(\phi, \psi)$ respectively, then it is obvious that in the lemma we are comparing stationary values. This version was proved by Smith in (68). Geometrically, this implies that if one saddle surface is above another then their stationary values must also satisfy this property. This is illustrated in figure (V.2.1).

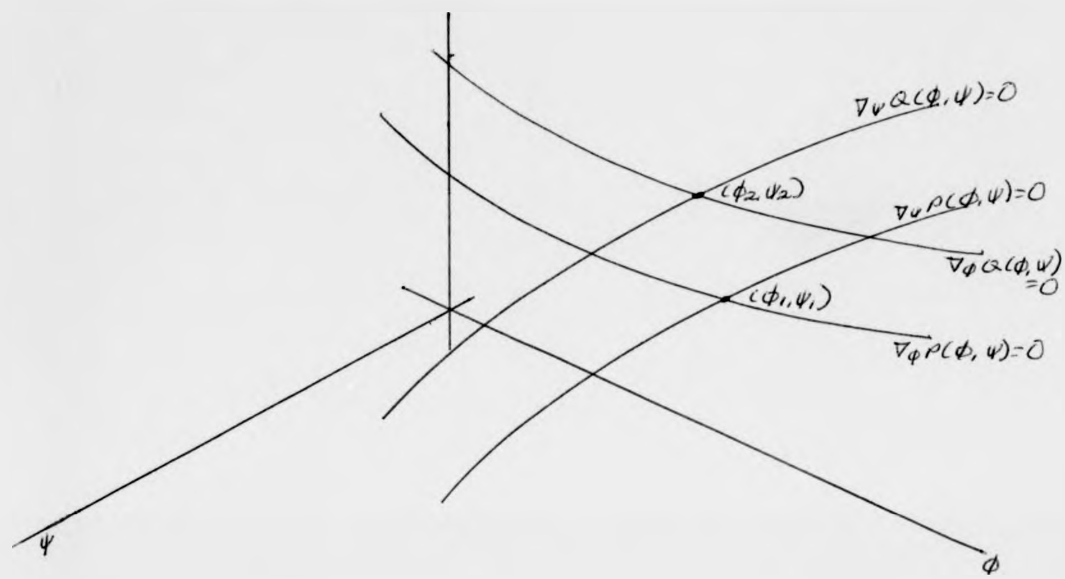


Figure (V.2.1)

Usually in applications, one of $P(\phi,\psi)$ and $Q(\phi,\psi)$ will be $L(\phi,\psi)$, the functional for which we want to bound the stationary value, and the other will be an approximation to $L(\phi,\psi)$, with a simpler structure. To obtain useful approximations between functionals, we require that the functionals are close in some way: one method of ensuring this is by requiring that

$$\inf_{(\phi,\psi)\in E^1\times F^1} \{Q(\phi,\psi) - P(\phi,\psi)\} \text{ is small.}$$

If this is zero, there is contact between the functionals at at least one point; the following two theorems exploit this property.

Theorem (V.2.1)

Let $L(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ be a differentiable convex/concave saddle functional with a stationary value at (ϕ_e,ψ_e), so that

$$\nabla_\phi L(\phi_e,\psi_e) = \nabla_\psi L(\phi_e,\psi_e) = 0 \qquad \text{(V.2.12)}$$

Suppose that we can find a functional $L_q(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ which is differentiable with respect to ϕ, and a pair $(\phi_o, \psi_o) \in E^1 \times F^1$, with the following properties:

(i) $L_a(\phi, \psi)$ is convex in ϕ

(ii) $W(\phi, \psi) = L_a(\phi, \psi) - L(\phi, \psi)$ is concave in ϕ

(iii) There exists at least one element $\phi_1 \in E^1$ such that

$$\nabla_\phi L_a(\phi_1, \psi_o) = \nabla_\phi W(\phi_o, \psi_o) \tag{V.2.13}$$

Then, if we let

$$J(\phi, \psi) = L_a(\phi, \psi) - W(\phi_o, \psi) + \langle \phi_o - \phi, \nabla_\phi W(\phi_o, \psi) \rangle \tag{V.2.14}$$

$$L(\phi_e, \psi_e) \geqslant J(\phi_1, \psi_o) \tag{V.2.15}$$

Proof

As $J(\phi, \psi) - L_a(\phi, \psi)$ is a functional linear in ϕ, $J(\phi, \psi)$ must be convex in ϕ for each fixed ψ.

Now, from equation (V.2.14),

$$\nabla_\phi J(\phi, \psi) = \nabla_\phi L_a(\phi, \psi) - \nabla_\phi W(\phi, \psi)$$

hence $\nabla_\phi J(\phi_1, \psi_o) = \nabla_\phi L_a(\phi_1, \psi_o) - \nabla_\phi W(\phi_o, \psi_o)$

$= 0$ from equation (V.2.13).

Equation (V.2.15) can be rewritten as

$$J(\phi, \psi) - L(\phi, \psi) = -\{W(\phi_o, \psi) - W(\phi, \psi) - \langle \phi_o - \phi, \nabla_\phi W(\phi_o, \psi) \rangle\}$$

$\leqslant 0$ as $W(\phi, \psi)$ is concave in ϕ

$$\text{hence } L(\phi, \psi) \geqslant J(\phi, \psi) \tag{V.2.16}$$

Therefore using lemma (V.2.1) with $L(\phi, \psi) = G(\phi, \psi)$ and $J(\phi, \psi) = P(\phi, \psi)$,

$L(\phi_e, \psi_e) \geqslant J(\phi_1, \psi_o)$.

From equation (V.2.14) we have

$$J(\phi_o, \psi) = L(\phi_o, \psi) \tag{V.2.17}$$

and so the two functionals $L(\phi, \psi)$ and $J(\phi, \psi)$ have contact along $\phi = \phi_o$

for all ψ . If $L_a(\phi, \psi) = L(\phi, \psi)$, $J(\phi, \psi) = L(\phi, \psi)$ and the lower bound reduces to the usual classical lower bound. If $J(\phi, \psi)$ is also concave in ψ , that is it is a convex/concave saddle functional, then, by the remarks following Lemma (V.2.1), (ϕ_1, ψ_o) is the stationary point of $J(\phi, \psi)$.

The counterpart of theorem (V.2.1) for upper bounds is given below in theorem (V.2.2).

Theorem (V.2.2)

Let $L(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ be a differentiable convex/concave saddle functional with a stationary value at (ϕ_e, ψ_e) so that

$$\nabla_\phi L(\phi_e, \psi_e) = \nabla_\psi L(\phi_e, \psi_e) = 0 \tag{V.2.18}$$

Assume that we can find a functional $L_b(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ which is differentiable with respect to ϕ , and a pair $(\phi_o, \psi_o) \in E^1 \times F^1$, with the following properties:

(i) $L_b(\phi, \psi)$ is concave in ψ

(ii) $V(\phi, \psi) = L_b(\phi, \psi) - L(\phi, \psi)$ is convex in ψ

(iii) There exists at least one element $\psi_1 \in F^1$ such that

$$\nabla_\psi L_b(\phi_o, \psi_1) = \nabla_\psi V(\phi_o, \psi_o) \tag{V.2.19}$$

Then, if we let

$$K(\phi, \psi) = L_b(\phi, \psi) - V(\phi, \psi_o) + \langle \psi_o - \psi_1, \nabla_\psi V(\phi, \psi_o) \rangle \tag{V.2.20}$$

$$L(\phi_e, \psi_e) \leq K(\phi_o, \psi_1) \tag{V.2.21}$$

Proof

As $K(\phi, \psi) - L_b(\phi, \psi)$ is linear in ψ , $K(\phi, \psi)$ must be concave in ψ for each fixed ϕ .

$$\nabla_\psi K(\phi, \psi) = \nabla_\psi L_b(\phi, \psi) - \nabla_\psi V(\phi, \psi_o)$$

and so $\nabla_\psi K(\phi_o, \psi_1) = \nabla_\psi L_b(\phi_o, \psi_1) - \nabla_\psi V(\phi_o, \psi_o)$

$= 0$ from equation (V.2.19)

We can write equation (V.2.20) as

$$K(\phi,\psi) - L(\phi,\psi) = V(\phi,\psi) - V(\phi,\psi_0) - \langle \psi - \psi_0, \nabla_\psi V(\phi,\psi_0)\rangle$$

$$\geq 0 \text{ as } V(\phi,\psi) \text{ is convex in } \psi$$

Hence $L(\phi,\psi) \leq K(\phi,\psi)$ (V.2.22)

Therefore using lemma (V.2.1) with $L(\phi,\psi) = P(\phi,\psi)$ and

$K(\phi,\psi) = Q(\phi,\psi)$ gives

$L(\phi_e,\psi_e) \leq K(\phi_0,\psi_1)$

If $K(\phi,\psi)$ is also convex in ϕ then (ϕ_0,ψ_1) is the stationary point of $K(\phi,\psi)$. If $L_b(\phi,\psi) = L(\phi,\psi)$, the upper bound reduces to the usual classical upper bound.

Before proving two more theorems involving comparison functionals which are different from theorems (V.2.1) and (V.2.2), we will show that these two theorems give the same bounds as those obtained in the main decomposition of functionals theorem, theorem (IV.2.2).

Theorem (IV.2.2) states:

If $M(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ and $N(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ are convex/concave saddle functionals and $L(\phi,\psi) = M(\phi,\psi) + N(\phi,\psi)$ has a stationary value (ϕ_e, ψ_e) given by

$\nabla_\phi L(\phi_e,\psi_e) = 0$ and $\nabla_\psi L(\phi_e,\psi_e) = 0$, then

$$M(\phi_2,\psi_2) + N(\phi_4,\psi_2) - \langle \phi_2 - \phi_4, \nabla_\phi M(\phi_2,\psi_2)\rangle$$

$$\leq L(\phi_e,\psi_e) \leq$$

$$M(\phi_1,\psi_1) + N(\phi_1,\psi_3) - \langle \psi_1 - \psi_3, \nabla_\psi M(\phi_1,\psi_1)\rangle \qquad (V.2.23)$$

where $\nabla_\phi M(\phi_2,\psi_2) + \nabla_\phi N(\phi_4,\psi_2) = 0$ (V.2.24)

and $\nabla_\psi M(\phi_1,\psi_1) + \nabla_\psi N(\phi_1,\psi_3) = 0$ (V.2.25)

We will need to deal with the lower and upper comparison bounds separately.

We can write equation (V.2.20) as

$$K(\phi,\psi) - L(\phi,\psi) = V(\phi,\psi) - V(\phi,\psi_o) - \langle \psi - \psi_o, \nabla_\psi V(\phi,\psi_o)\rangle$$

$$\geq 0 \quad \text{as } V(\phi,\psi) \text{ is convex in } \psi$$

Hence $L(\phi,\psi) \leq K(\phi,\psi)$ (V.2.22)

Therefore using lemma (V.2.1) with $L(\phi,\psi) = P(\phi,\psi)$ and

$K(\phi,\psi) = Q(\phi,\psi)$ gives

$L(\phi_e,\psi_e) \leq K(\phi_o,\psi_1)$

If $K(\phi,\psi)$ is also convex in ϕ then (ϕ_o,ψ_1) is the stationary point of $K(\phi,\psi)$. If $L_b(\phi,\psi) = L(\phi,\psi)$, the upper bound reduces to the usual classical upper bound.

Before proving two more theorems involving comparison functionals which are different from theorems (V.2.1) and (V.2.2), we will show that these two theorems give the same bounds as those obtained in the main decomposition of functionals theorem, theorem (IV.2.2).

Theorem (IV.2.2) states:

If $M(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ and $N(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ are convex/concave saddle functionals and $L(\phi,\psi) = M(\phi,\psi) + N(\phi,\psi)$ has a stationary value (ϕ_e,ψ_e) given by

$\nabla_\phi L(\phi_e,\psi_e) = 0$ and $\nabla_\psi L(\phi_e,\psi_e) = 0$, then

$$M(\phi_2,\psi_2) + N(\phi_4,\psi_2) - \langle \phi_2 - \phi_4, \nabla_\phi M(\phi_2,\psi_2)\rangle$$

$$\leq L(\phi_e,\psi_e) \leq$$

$$M(\phi_1,\psi_1) + N(\phi_1,\psi_3) - \langle \psi_1 - \psi_3, \nabla_\psi M(\phi_1,\psi_1)\rangle \qquad \text{(V.2.23)}$$

where $\nabla_\phi M(\phi_2,\psi_2) + \nabla_\phi N(\phi_4,\psi_2) = 0$ (V.2.24)

and $\nabla_\psi M(\phi_1,\psi_1) + \nabla_\psi N(\phi_1,\psi_3) = 0$ (V.2.25)

We will need to deal with the lower and upper comparison bounds separately.

(i) Lower Bound

Essentially, theorem (V.2.1) states:

If $L_a(\phi,\psi)$ and $L(\phi,\psi) - L_a(\phi,\psi)$ are both convex in ϕ ,

$$L_a(\phi_1,\psi_o) + \{L(\phi_o,\psi_o) - L_a(\phi_o,\psi_o)\}$$
$$-\langle\phi_1 - \phi_o,\ \nabla_\phi L_a(\phi_1,\psi_o)\rangle \le L(\phi_e,\psi_e) \qquad (V.2.26)$$

where $\nabla_\phi L_a(\phi_1,\psi_o) + \nabla_\phi L_a(\phi_o,\psi_o) - \nabla_\phi L(\phi_o,\psi_o) = 0$ (V.2.27)

Let the functionals and trial functions in the above equations be changed as follows:

$L_a(\phi,\psi) = M(\phi,\psi)$ $\quad L(\phi,\psi) - L_a(\phi,\psi) = N(\phi,\psi)$

$\phi_1 = \phi_2$, $\quad \phi_o = \phi_4$ and $\psi_o = \psi_2$

Then theorem (V.2.1) becomes:

If $M(\phi,\psi)$ and $N(\phi,\psi)$ are both convex in ϕ ,

$$M(\phi_2,\psi_2) + N(\phi_4,\psi_2) - \langle\phi_2 - \phi_4,\ \nabla_\phi M(\phi_2,\psi_2)\rangle$$
$$\le L(\phi_e,\psi_e) \qquad (V.2.28)$$

where $\nabla_\phi M(\phi_2,\psi_2) + \nabla_\phi N(\phi_4,\psi_2) = 0$ (V.2.29)

It can clearly be seen, by comparing equations (V.2.23) and (V.2.24) with (V.2.28) and (V.2.29) respectively, that the lower bound in theorem (IV.2.2) is the same as the bound in theorem (V.2.1); the only difference is that in theorem (IV.2.2) we require $M(\phi,\psi)$ and $N(\phi,\psi)$ to be concave in ψ , with at least one strictly concave, whereas in theorem (V.2.1) we do not make this requirement. However, as $L(\phi,\psi)$ is strictly concave in ψ , and $L_a(\phi,\psi) + \{L(\phi,\psi) - L_a(\phi,\psi)\} = L(\phi,\psi)$, both $L_a(\phi,\psi)$ and $L(\phi,\psi) - L_a(\phi,\psi)$ must be concave in ψ , with at least one strictly concave; hence the decomposition lower bound and comparison bound are identical.

(ii) Upper Bound

Theorem (V.2.2) can be written:

If $L_b(\phi,\psi)$ and $L(\phi,\psi) - L_b(\phi,\psi)$ are both concave in ψ ,

$$L_b(\phi_o, \psi_1) + L(\phi_o, \psi_o) - L_a(\phi_o, \psi_o)$$

$$- \langle \psi_1 - \psi_o, \nabla_\psi L_b(\phi_o, \psi_1) \rangle \geqslant L(\phi_e, \psi_e) \qquad (V.2.30)$$

where $\nabla_\psi L_b(\phi_o, \psi_1) + \nabla_\psi L(\phi_o, \psi_o) - \nabla_\psi L_b(\phi_o, \psi_o) = 0$ (V.2.31)

Choose the functionals and trial functions in the above equations as follows:

$L_b(\phi, \psi) = M(\phi, \psi)$, $L(\phi, \psi) - L_b(\phi, \psi) = N(\phi, \psi)$

$\phi_o = \phi_1$, $\psi_1 = \psi_1$, $\psi_o = \psi_3$

Then theorem (V.2.2) becomes:

If $M(\phi, \psi)$ and $N(\phi, \psi)$ are both concave in ψ ,

$$M(\phi_1, \psi_1) + N(\phi_1, \psi_3) - \langle \psi_1 - \psi_3, \nabla_\psi M(\phi_1, \psi_1) \rangle$$

$$\geqslant L(\phi_e, \psi_e) \qquad (V.2.32)$$

where $\nabla_\psi M(\phi_1, \psi_1) + \nabla_\psi N(\psi_1, \psi_3) = 0$ (V.2.33)

Again, it is obvious, by comparing equations (V.2.23) and (V.2.25) with (V.2.32) and (V.2.33) respectively, that the upper bound in theorem (IV.2.2) is the same as the bound in theorem (V.2.2); As $L_b(\phi, \psi) + \{L(\phi, \psi) - L_b(\phi, \psi)\} = L(\phi, \psi)$, both $L_b(\phi, \psi)$ and $L(\phi, \psi) - L_b(\phi, \psi)$ must be convex in ϕ , with at least one strictly convex, and therefore the decomposition upper bound and comparison bound are identical.

It is interesting to see that the same bounds can be obtained from different assumptions and from a different angle. As the decomposition bounds in theorem (IV.2.2) were dealt with extensively in the previous chapter, the bounds obtained in theorems (V.2.1) and (V.2.2) will not be considered in the rest of this chapter.

Alternative lower and upper bounds to those given in theorems (V.2.1) and (V.2.2) can be constructed if we assume that the 'approximate' functionals $L_a(\phi, \psi)$ and $L_b(\phi, \psi)$ are both convex/concave saddle functionals, and if the 'difference' functionals $W(\phi,\psi)$ and $V(\phi,\psi)$ are restricted further; theorems (V.2.3) and (V.2.4) provide these bounds.

Theorem (V.2.3)

Let $L(\phi,\psi) : E^1 \times F^1 \to I\!R$ be a differentiable convex/concave saddle functional with a stationary value at (ϕ_e, ψ_e). Assume that there exists a differentiable convex/concave saddle functional $L_a(\phi,\psi) : E^1 \times F^1 \to I\!R$, and a pair $(\phi_o, \psi_o) \in E^1 \times F^1$ with the following properties:

(i) $W(\phi,\psi) = L_a(\phi,\psi) - L(\phi,\psi)$ is jointly concave in ϕ and ψ

(ii) There exists a pair $(\phi_1, \psi_1) \in E^1 \times F^1$ such that

$$\nabla_\phi L_a(\phi_1, \psi_1) = \nabla_\phi W(\phi_o, \psi_o) \tag{V.2.34}$$

and

$$\nabla_\psi L_a(\phi_1, \psi_1) = \nabla_\psi W(\phi_o, \psi_o) \tag{V.2.35}$$

Then, if we define $J(\phi,\psi)$ by

$$J(\phi,\psi) = L_a(\phi,\psi) - W(\phi_o, \psi_o) + \langle \phi_o - \phi, \nabla_\phi W(\phi_o,\psi_o) \rangle + \langle \psi_o - \psi, \nabla_\psi W(\phi_o,\psi_o) \rangle \tag{V.2.36}$$

we have $L(\phi_e, \psi_e) \geqslant J(\phi_1, \psi_1)$ (V.2.37)

Proof

$J(\phi,\psi) - L_a(\phi,\psi)$ is linear in both ϕ and ψ and hence as $L_a(\phi,\psi)$ is a convex/concave saddle functional, so is $J(\phi,\psi)$.

From equation (V.2.36),

$$\nabla_\phi J(\phi,\psi) = \nabla_\phi L_a(\phi,\psi) - \nabla_\phi W(\phi_o, \psi_o) \tag{V.2.38}$$

and

$$\nabla_\psi J(\phi,\psi) = \nabla_\psi L_a(\phi,\psi) - \nabla_\psi W(\phi_o, \phi_o) \tag{V.2.39}$$

The stationary value of $J(\phi,\psi)$ occurs when $\nabla_\phi J(\phi,\psi) = \nabla_\psi J(\phi,\psi) = 0$; comparing equation (V.2.34) with (V.2.38) and (V.2.35) with (V.2.39), we can see that $J(\phi,\psi)$ has its stationary value at (ϕ_1, ψ_1).

Using equation (V.2.36),

$$J(\phi,\psi) - L(\phi,\psi) = -\{ - W(\phi,\psi) + W(\phi_o,\psi_o)$$
$$+\langle \phi - \phi_o, \nabla_\phi W(\phi_o,\psi_o)\rangle$$
$$+\langle \psi - \psi_o, \nabla_\psi W(\phi_o,\psi_o)\rangle$$

$\leq$ 0 as $W(\phi,\psi)$ is jointly concave in ϕ and ψ (using equation (IV.2.6)).

Hence $L(\phi,\psi) \geq J(\phi,\psi)$ (V.2.40)

Therefore, by Lemma (V.2.1),

$$L(\phi_e,\psi_e) \geq J(\phi_1,\psi_1)$$

A theorem similar to the above was proved by Smith in (68), with the additional assumption that $\nabla\phi\ L(\phi_o,\psi_o) = 0$ making the pair (ϕ_o, ψ_o) trial functions for the usual classical dual extremum principles lower bound.

If $\nabla_\phi\ L(\phi_o,\psi_o) = 0$, and using equations (V.2.34) to (V.2.36),

$$J(\phi_1,\psi_1) - L(\phi_o,\psi_o) = L_a(\phi_1,\psi_1) - L_a(\phi_o,\psi_o)$$
$$-\langle \phi_1 - \phi_o, \nabla_\phi L_a(\phi_o,\psi_o)\rangle$$
$$-\langle \psi_1 - \psi_o, \nabla_\psi L_a(\phi_1,\psi_1)\rangle$$

$\geq$ 0 as $L_a(\phi,\psi)$ is a convex/concave saddle functional (see definition (II.9.2)).

Hence, as noted in paper (68), in this case

$$L(\phi_e,\psi_e) \geq J(\phi_1,\psi_1) \geq L(\phi_o,\psi_o)$$

thus giving an automatic improvement over the classical lower bound.

If $L_a(\phi,\psi) = L(\phi,\psi)$, $J(\phi,\psi) = L(\phi,\psi)$ and the lower bound reduces to the usual classical lower bound.

<u>Theorem (V.2.4)</u>

Let $L(\phi,\psi) : E^1 \times F^1 \to \mathbb{R}$ be a differentiable convex/concave saddle functional with a stationary value at (ϕ_e,ψ_e).

Assume that there exists a differentiable convex/concave saddle functional

$L_b(\phi, \psi)$ and a pair $(\phi_o, \psi_o) \in E^1 \times F^1$ with the following properties:

(i) $V(\phi, \psi) = L_b(\phi, \psi) - L(\phi, \psi)$ is jointly convex in ϕ and ψ

(ii) There exists a pair $(\phi_1, \psi_1) \in E^1 \times F^1$ such that

$$\nabla_\phi L_b(\phi_1, \psi_1) = \nabla_\phi V(\phi_o, \psi_o) \tag{V.2.41}$$

and

$$\nabla_\psi L_b(\phi_1, \psi_1) = \nabla_\psi V(\phi_o, \psi_o) \tag{V.2.42}$$

Then, if we define $K(\phi, \psi)$ by

$$K(\phi, \psi) = L_b(\phi, \psi) - V(\phi_o, \psi_o) + \langle \phi_o - \phi, \nabla_\phi V(\phi_o, \psi_o) \rangle + \langle \psi_o - \psi, \nabla_\psi V(\phi_o, \psi_o) \rangle \tag{V.2.43}$$

we have $L(\phi_e, \psi_e) \leq K(\phi_1, \psi_1)$ (V.2.44)

<u>Proof</u>

$K(\phi, \psi)$ is a convex/concave saddle functional as $K(\phi, \psi) - L_b(\phi, \psi)$ is linear in ϕ and ψ and $L_b(\phi, \psi)$ is a convex/concave saddle functional.

From equation (V.2.32),

$$\nabla_\phi K(\phi, \psi) = \nabla_\phi L_b(\phi, \psi) - \nabla_\phi V(\phi_o, \psi_o) \tag{V.2.45}$$

and $\nabla_\psi K(\phi, \psi) = \nabla_\psi L_b(\phi, \psi) - \nabla_\psi V(\phi_o, \psi_o)$ (V.2.46)

and hence comparing equation (V.2.41) with (V.2.45) and (V.2.42) with (V.2.46), $K(\phi, \psi)$ has a stationary value at (ϕ_1, ψ_1).

From equation (V.2.43),

$$\begin{aligned} K(\phi, \psi) - L(\phi, \psi) &= - V(\phi_o, \psi_o) + V(\phi, \psi) \\ &\quad + \langle \phi_o - \phi_1, \nabla_\phi V(\phi_o, \psi_o) \rangle \\ &\quad + \langle \psi_o - \psi_1, \nabla_\psi V(\phi_o, \psi_o) \rangle \\ &\geqslant 0 \text{ by equation (V.2.5) as } V(\phi, \psi) \text{ is jointly convex in} \end{aligned}$$

ϕ and ψ ; hence $L(\phi, \psi) \leq K(\phi, \psi)$ (V.2.47)

and therefore, by Lemma (V.2.1), $L(\phi_e, \psi_e) \leq K(\phi_1, \psi_1)$.

If $L_b(\phi, \psi) = L(\phi, \psi)$, $K(\phi, \psi) = L(\phi, \psi)$ and the upper bound reduces to the usual classical upper bound.

CHAPTER V

V.3 Application to quadratic functional

We are going to apply theorems (V.2.3) and (V.2.4) to the usual quadratic convex/concave saddle functional given by the equation

$$L(\phi, \psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B\phi\rangle - \tfrac{1}{2}\langle\psi, C\psi\rangle + \langle\phi, f\rangle + \langle\psi, g\rangle \tag{V.3.1}$$

where A is a linear operator with an adjoint A^x such that

$$\langle\phi, A\psi\rangle = \langle\psi, A^x\phi\rangle \tag{V.3.2}$$

and B and C are linear, symmetric, positive-definite operators.

The functional derivatives of $L(\phi, \psi)$ are given by

$$\nabla_\phi L(\psi, \psi) = A\psi + B\phi + f \quad \text{and} \tag{V.3.3}$$

$$\nabla_\psi L(\phi, \psi) = A^x\phi - C\psi + g; \tag{V.3.4}$$

from section (II.12), the classical dual extremum principles are

$$L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha) \tag{V.3.5}$$

with $L(\phi_\beta, \psi_\beta) = -\tfrac{1}{2}\langle\phi_\beta, B\phi_\beta\rangle - \tfrac{1}{2}\langle\psi_\beta, C\psi_\beta\rangle + \langle\psi_\beta, g\rangle$ (V.3.6)

where $A\psi_\beta + B\phi_\beta + f = 0$, (V.3.7)

$$L(\phi_\alpha, \psi_\alpha) = \tfrac{1}{2}\langle\phi_\alpha, B\phi_\alpha\rangle + \tfrac{1}{2}\langle\psi_\alpha, C\psi_\alpha\rangle + \langle\phi_\alpha, f\rangle \tag{V.3.8}$$

where $A^x\phi_\alpha - C\psi_\alpha + g = 0$ (V.3.9)

and $L(\phi_e, \psi_e) = \tfrac{1}{2}\langle\phi_e, f\rangle + \tfrac{1}{2}\langle\psi_e, g\rangle$ (V.3.10)

where $A\psi_e + B\phi_e + f = 0$, and $A^x\phi_e - C\psi_e + g = 0$ (V.3.11)

To apply theorems (V.2.3) and (V.2.4) to the quadratic functional, we need convex/concave saddle functionals $L_a(\phi, \psi)$ and $L_b(\phi, \psi)$ which also contain the term $\langle\phi, A\psi\rangle$, as this needs to be eliminated from the difference functionals $W(\phi, \psi)$ and $V(\phi, \psi)$, as it is not jointly convex or jointly concave (see example 1 after definition (II.8.8)).

Theorem (V.2.3)

Let $L_a(\phi, \psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B_a\phi\rangle - \tfrac{1}{2}\langle\psi, C_a\psi\rangle + \langle\phi, f\rangle + \langle\psi, g\rangle$

(V.3.12)

where B_a and C_a are linear, symmetric, positive-definite operators to ensure that $L_a(\phi, \psi)$ is a convex/concave saddle functional. This gives the difference functional $W(\phi, \psi)$ as:

$$W(\phi, \psi) = L_a(\phi, \psi) - L(\phi, \psi)$$

$$= -\tfrac{1}{2}\langle \phi, (B - B_a)\phi\rangle - \tfrac{1}{2}\langle \psi, (C_a - C)\psi\rangle \qquad \text{(V.3.13)}$$

which is jointly concave in ϕ and ψ if $B - B_a$ and $C_a - C$ are positive operators - that is, we require

$$B \geqslant B_a > 0 \quad \text{and} \quad C_a \geqslant C > 0 \qquad \text{(V.3.14)}$$

We then have the comparison lower bound $J(\phi_1, \psi_1)$, where

$$\begin{aligned} J(\phi, \psi) = {} & \langle \phi, A\psi\rangle + \tfrac{1}{2}\langle \phi, B_a\phi\rangle - \tfrac{1}{2}\langle \psi, C_a\psi\rangle \\ & + \langle \phi, f\rangle + \langle \psi, g\rangle \\ & + \langle \phi, (B - B_a)\phi_o\rangle + \langle \psi, (C_a - C)\psi_o\rangle \\ & - \tfrac{1}{2}\langle \phi_o, (B - B_a)\phi_o\rangle - \tfrac{1}{2}\langle \psi_o, (C_a - C)\psi_o\rangle \end{aligned} \qquad \text{(V.3.15)}$$

and $$A\psi_1 + B_a\phi_1 + f = (B_a - B)\phi_o, \qquad \text{(V.3.16)}$$

$$A^x\phi_1 - C_a\psi_1 + g = (C - C_a)\psi_o \qquad \text{(V.3.17)}$$

Using these equations,

$$\begin{aligned} J(\phi_1, \psi_1) = {} & -\tfrac{1}{2}\langle \phi_1, B\phi_1\rangle - \tfrac{1}{2}\langle \psi_1, C\psi_1\rangle + \langle \psi_1, g\rangle \\ & + \tfrac{1}{2}\langle \phi_1, (B - B_a)\phi_1\rangle - \tfrac{1}{2}\langle \phi_o, (B - B_a)\phi_o\rangle \\ & - \tfrac{1}{2}\langle \psi_o - \psi_1, (C_a - C)(\psi_o - \psi_1)\rangle \end{aligned} \qquad \text{(V.3.18)}$$

From the last three equations, the comparison lower bound reduces to the classical lower bound if either $B = B_a$ and $C = C_a$, $\phi_1 = \phi_o$ and $\psi_1 = \psi_o$, $B = B_a$ and $\psi_o = \psi_1$ or $C = C_a$ and $\phi_o = \phi_1$

<u>Theorem (V.2.4)</u>

Let $$\begin{aligned} L_b(\phi, \psi) = {} & \langle \phi, A\psi\rangle + \tfrac{1}{2}\langle \phi, B_b\phi\rangle - \tfrac{1}{2}\langle \psi, C_b\psi\rangle \\ & + \langle \phi, f\rangle + \langle \psi, g\rangle \end{aligned} \qquad \text{(V.3.19)}$$

where B_b and C_b are linear, symmetric, positive-definite operators; then $L_b(\phi, \psi)$ is a convex/concave saddle functional.

Hence $V(\phi, \psi) = L_b(\phi, \psi) - L(\phi, \psi)$ becomes

$$V(\phi,\psi) = \tfrac{1}{2}\langle \phi, (B_b - B)\phi\rangle + \tfrac{1}{2}\langle \psi, (C - C_b)\psi\rangle \qquad (V.3.20)$$

which is jointly convex in ϕ and ψ if $B_b - B$ and $C - C_b$ are positive operators; we therefore require

$$B_b \geqslant B > 0 \quad \text{and} \quad C \geqslant C_b > 0 \qquad (V.3.21)$$

The comparison upper bound is then $K(\phi_1, \psi_1)$, where

$$\begin{aligned} K(\phi,\psi) = {} & \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle \phi, B_b\phi\rangle - \tfrac{1}{2}\langle \psi, C_b\psi\rangle \\ & + \langle\phi, f\rangle + \langle\psi, g\rangle \\ & + \tfrac{1}{2}\langle \phi_o, (B_b - B)\phi_o\rangle + \tfrac{1}{2}\langle \psi_o, (C - C_b)\psi_o\rangle \\ & - \langle \phi, (B_b - B)\phi_o\rangle - \langle\psi, (C - C_b)\psi_o\rangle \qquad (V.3.22) \end{aligned}$$

and $A\psi_1 + B_b\phi_1 + f = (B_b - B)\phi_o$, $\qquad (V.3.23)$

$$A^x\phi_1 - C_b\psi_1 + g = (C - C_b)\phi_o \qquad (V.3.24)$$

These equations give

$$\begin{aligned} K(\phi_1, \psi_1) = {} & \tfrac{1}{2}\langle\phi_1, B\phi_1\rangle + \tfrac{1}{2}\langle\psi_1, C\psi_1\rangle + \langle\phi_1, f\rangle + \tfrac{1}{2}\langle\phi_o - \phi_1, \\ & \qquad (B_b - B)(\phi_o - \phi_1)\rangle \\ & + \tfrac{1}{2}\langle\psi_o, (C - C_b)\psi_o\rangle - \tfrac{1}{2}\langle\psi_1, (C - C_b)\psi_1\rangle \qquad (V.3.25) \end{aligned}$$

From the last three equations, the comparison upper bound reduces to the classical upper bound if either $B = B_b$ and $C = C_b$, $\phi_o = \phi_1$ and $\psi_o = \psi_1$, $B = B_b$ and $\psi_o = \psi_1$, or $C = C_b$ and $\phi_o = \phi_1$.

As in the previous chapter, we want to find conditions on the comparison operators and/or the trial functions in the comparison bounds which ensure in each case that the comparison bounds are better than the classical bounds. We do this by first finding the difference between each comparable pair of classical and comparison bounds.

<u>Theorem (V.2.3)</u>

The classical lower bound is

$$L(\phi_s, \psi_s) = -\tfrac{1}{2}\langle\phi_s, B\phi_s\rangle - \tfrac{1}{2}\langle\psi_s, C\psi_s\rangle + \langle\psi_s, g\rangle \qquad (V.3.26)$$

where $A\psi_s + B\phi_s + f = 0 \qquad (V.3.27)$

and the comparison lower bound from equations (V.3.16) - (V.3.18) is

$$J(\phi_1, \psi_1) = -\tfrac{1}{2}\langle \phi_1, B\phi_1\rangle - \tfrac{1}{2}\langle \psi_1, C\psi_1\rangle + \langle \psi_1, g\rangle + \tfrac{1}{2}\langle \phi_1, (B - B_a)\phi_1\rangle - \tfrac{1}{2}\langle \phi_0, (B - B_a)\phi_0\rangle - \tfrac{1}{2}\langle \psi_0 - \psi_1, (C_a - C)(\psi_0 - \psi_1)\rangle \quad \text{(V.3.28)}$$

where $A\psi_1 + B\phi_0 + f = B_a(\phi_0 - \phi_1)$ (V.3.29)

and $A^x\phi_1 - C\psi_0 + g = C_a(\psi_0 - \psi_1)$ (V.3.30)

$J(\phi_1, \psi_1)$ will be a better bound than $L(\phi_s, \psi_s)$ if

$J(\phi_1, \psi_1) - L(\phi_s, \psi_s) > 0$ (V.3.31)

From equations (V.3.26) to (V.3.30),

$$J(\phi_1, \psi_1) - L(\phi_s, \psi_s) = -\tfrac{1}{2}\langle \phi_1, B\phi_1\rangle + \tfrac{1}{2}\langle \phi_s, B\phi_s\rangle - \tfrac{1}{2}\langle \psi_1, C\psi_1\rangle + \tfrac{1}{2}\langle \psi_s, C\psi_s\rangle + \langle \psi_1 - \psi_s, g\rangle - \tfrac{1}{2}\langle \psi_0 - \psi_1, (C_a - C)(\psi_0 - \psi_1)\rangle + \tfrac{1}{2}\langle \phi_1, (B - B_a)\phi_1\rangle - \tfrac{1}{2}\langle \phi_0, (B - B_a)\phi_0\rangle \quad \text{(V.3.32)}$$

where $A(\psi_1 - \psi_s) + B(\phi_0 - \phi_s) - B_a(\phi_0 - \phi_1) = 0$

and $A^x\phi_1 - C\psi_1 + g + (C_a - C)(\psi_0 - \psi_1) = 0$ (V.3.33)

Using the last two equations,

$$J(\phi_1, \psi_1) - L(\phi_s, \psi_s) = \tfrac{1}{2}\langle \phi_1 - \phi_s, B(\phi_1 - \phi_s)\rangle + \tfrac{1}{2}\langle \psi_1 - \psi_s, C(\psi_1 - \psi_s)\rangle + \tfrac{1}{2}\langle \psi_0 - \psi_1, (C_a - C)(\psi_0 - \psi_1)\rangle - \tfrac{1}{2}\langle \phi_0 - \phi_1, (B - B_a)(\phi_0 - \phi_1)\rangle + \langle \psi_0 - \psi_s, A^x\phi_1 - C\psi_1 + g\rangle \quad \text{(V.3.34)}$$

As B and C are positive-definite operators, and $(C_a - C)$ and $(B - B_a)$ are positive operators, it is clear from equation (V.3.34) that $J(\phi_1, \psi_1) - L(\phi_s, \psi_s)$ is greater than zero if $B = B_a$ and $\psi_0 = \psi_s$ (V.3.35)

Hence $J(\phi_1, \psi_1) = -\tfrac{1}{2}\langle \phi_1, B\phi_1\rangle - \tfrac{1}{2}\langle \psi_1, C\psi_1\rangle + \langle \psi_1, g\rangle - \tfrac{1}{2}\langle \psi_0 - \psi_1, (C_a - C)(\psi_0 - \psi_1)\rangle$ (V.3.36) where $A\psi_1 + B\phi_1 + f = 0$ (V.3.37)

and $A^x\phi_1 - C\psi_0 + g + C_a(\psi_0 - \psi_1) = 0$ (V.3.38)

will be a sharper bound than the classical bound given by equations (V.3.26) and (V.3.27) if ψ_0 in equation (V.3.38) is chosen equal to ψ_α in equation (V.3.27).

Theorem (V.2.4)

The classical upper bound is

$$L(\phi_\alpha, \psi_\alpha) = \tfrac{1}{2}\langle \phi_\alpha , B\phi_\alpha\rangle + \tfrac{1}{2}\langle \psi_\alpha, C\psi_\alpha\rangle + \langle \phi_\alpha, f\rangle \qquad \text{(V.3.39)}$$

where $A^x\phi_\alpha - C\psi_\alpha + g = 0$ (V.3.40)

and the comparison upper bound from equations (V.3.23) to (V.3.25) is

$$\begin{aligned} K(\phi_1,\psi_1) = {} & \tfrac{1}{2}\langle\phi_1, B\phi_1\rangle + \tfrac{1}{2}\langle\psi_1, C\psi_1\rangle + \langle\phi_1, f\rangle \\ & + \tfrac{1}{2}\langle \phi_0 - \phi_1, (B_b - B)(\phi_0 - \phi_1)\rangle \\ & + \tfrac{1}{2}\langle \psi_0, (C - C_b)\psi_0\rangle - \tfrac{1}{2}\langle\psi_1, (C - C_b)\psi_1\rangle \end{aligned} \qquad \text{(V.3.41)}$$

where $A\psi_1 + B_b\phi_1 + f - (B_b - B)\phi_0 = 0$ (V.3.42)

and $A^x\phi_1 - C_b\psi_1 + g - (C - C_b)\psi_0 = 0$ (V.3.43)

$K(\phi_1, \psi_1)$ will be a better bound than $L(\phi_\alpha, \psi_\alpha)$ if

$$L(\phi_\alpha, \psi_\alpha) - K(\phi_1,\psi_1) > 0 \qquad \text{(V.3.44)}$$

From equations (V.3.39) to (V.3.43)

$$\begin{aligned} L(\phi_\alpha,\psi_\alpha) - K(\phi_1,\psi_1) = {} & - \tfrac{1}{2}\langle\phi_1, B\phi_1\rangle + \tfrac{1}{2}\langle\phi_\alpha, B\phi_\alpha\rangle \\ & - \tfrac{1}{2}\langle\psi_1, C\psi_1\rangle + \tfrac{1}{2}\langle\psi_\alpha, C\psi_\alpha\rangle \\ & + \langle\phi_\alpha - \phi_1, f\rangle \\ & - \tfrac{1}{2}\langle\phi_0 - \phi_1, (B_b - B)(\phi_0 - \phi_1)\rangle \\ & - \tfrac{1}{2}\langle\psi_0, (C - C_b)\psi_0\rangle \\ & + \tfrac{1}{2}\langle\psi_1, (C - C_b)\psi_1\rangle \end{aligned} \qquad \text{(V.3.45)}$$

where $A^x(\phi_1 - \phi_\alpha) - C(\psi_0 - \psi_\alpha) + C_b(\psi_0 - \psi_1) = 0;$ (V.3.46)

also $A\psi_1 + B\phi_1 + f + (B - B_b)(\phi_0 - \phi_1) = 0$ (V.3.47)

Using the last two equations,

$$\begin{aligned} L(\phi_\alpha, \psi_\alpha) - K(\phi_1, \psi_1) &= \tfrac{1}{2}\langle \phi_1 - \phi_\alpha\,,\ B(\phi_1 - \phi_\alpha)\rangle \\ &+ \tfrac{1}{2}\langle \psi_1 - \psi_\alpha\,,\ C(\psi_1 - \psi_\alpha)\rangle \\ &+ \tfrac{1}{2}\langle \phi_o - \phi_1,\ (B_b - B)\ (\phi_o - \phi_1)\rangle \\ &- \tfrac{1}{2}\langle \psi_o - \psi_1,\ (C - C_b)\ (\psi_o - \psi_1)\rangle \\ &- \langle \phi_o - \phi_\alpha\,,\ A\psi_1 + B\phi_1 + f\rangle \end{aligned} \qquad (V.3.48)$$

As B and C are positive-definite operators, and B_b - B, C - C_b are positive operators, it is obvious from the last equation that $L(\phi_\alpha, \psi_\alpha) - K(\phi_1, \psi_1)$ is greater than zero if $\phi_o = \phi_\alpha$, and $C = C_b$ (V.3.49)

Hence $$K(\phi_1, \psi_1) = \tfrac{1}{2}\langle \phi_1, B\phi_1\rangle + \tfrac{1}{2}\langle \psi_1, C\psi_1\rangle + \langle \phi_1, f\rangle + \tfrac{1}{2}\langle \phi_o - \phi_1, (B_b - B)\ (\phi_o - \phi_1)\rangle \qquad (V.3.50)$$

where $A\psi_1 + B\phi_o + f - B_b(\phi_o - \phi_1) = 0$ (V.3.51)

and $A^x\phi_1 - C\psi_1 + g = 0$ (V.3.52)

will be a sharper bound than the classical bound given by equations (V.3.39) and (V.3.40) of ϕ_o in equation (V.3.51) is chosen equal to ϕ_α in equation (V.3.40).

Obviously, there is no guarantee that the lower bounds $L(\phi_\beta, \psi_\beta)$ and $J(\phi_1, \psi_1)$ or upper bounds $L(\phi_\alpha, \psi_\alpha)$ and $K(\phi_1, \psi_1)$ are close to $L(\phi_e, \psi_e)$. The combination of iterative methods with the comparison bounds, as described in the next section, could be used to obtain sharp bounds.

CHAPTER V

V.4 Comparison Functionals and Iterative Schemes

In this section we intend to set up iterative schemes involving the comparison bounds for the general quadratic functional using theorems (V.2.3) and (V.2.4). We shall only consider cobweb iterative schemes; ideally we would have liked to find schemes similar to the optimising schemes in chapter III, with $J(\phi_1, \psi_1)$ from equation (V.3.18) and $K(\phi_1, \psi_1)$ from equation (V.3.25), as the bounds to be optimised; but as these bounds contain the extra trial functions ϕ_0 and ψ_0, this has not been possible. Note however, that an optimising iterative scheme for the bounds in theorems (V.2.1) and (V.2.2) has been discussed in section (IV.5), when the method was applied to the equivalent decomposition bounds.

Convergence conditions for the iterative schemes are found; after summarising the schemes together with their convergence conditions, examples are provided to illustrate that there are problems for which the conditions can be satisfied.

The section ends with a brief discussion concerning iterative schemes for which we cannot prove convergence.

Theorem (V.2.3)

Let $L(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ be defined by

$$L(\phi, \psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B\phi\rangle - \tfrac{1}{2}\langle\psi, C\psi\rangle + \langle\phi, f\rangle + \langle\psi, g\rangle \tag{V.4.1}$$

From the last section, we define $L_a(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$ by the equation

$$L_a(\phi, \psi) = \langle\phi, A\psi\rangle + \tfrac{1}{2}\langle\phi, B_a\phi\rangle - \tfrac{1}{2}\langle\psi, C_a\psi\rangle + \langle\phi, f\rangle + \langle\psi, g\rangle \tag{V.4.2}$$

where B_a and C_a are linear, symmetric, positive-definite equations such that

$0 < B_a \leq B$ and $0 < C \leq C_a$

Then the comparison lower bound $J(\phi_1, \psi_1)$ is given by

$$J(\phi_1, \psi_1) = -\tfrac{1}{2}\langle\phi_1, B\phi_1\rangle - \tfrac{1}{2}\langle\psi_1, C\psi_1\rangle + \langle\psi_1, g\rangle$$
$$+ \tfrac{1}{2}\langle\phi_1, (B - B_a)\phi_1\rangle - \tfrac{1}{2}\langle\phi_o, (B - B_a)\phi_o\rangle$$
$$- \tfrac{1}{2}\langle\psi_o - \psi_1, (C_a - C)(\psi_o - \psi_1)\rangle \qquad (V.4.4)$$

where $A\psi_1 + B_a\phi_1 + f = (B_a - B)\phi_o$ (V.4.5)

and $A^x\phi_1 - C_a\psi_1 + g = (C - C_a)\psi_o$ (V.4.6)

We can set up a cobweb iterative scheme by letting $(\phi_o, \psi_o) = (\phi_{n-1}, \psi_{n-1})$ and $(\phi_1, \psi_1) = (\phi_n, \psi_n)$; this results in

$$J(\phi_n, \psi_n) = -\tfrac{1}{2}\langle\phi_n, B\phi_n\rangle - \tfrac{1}{2}\langle\psi_n, C\psi_n\rangle + \langle\psi_n, g\rangle$$
$$+ \tfrac{1}{2}\langle\phi_n, (B - B_a)\phi_n\rangle - \tfrac{1}{2}\langle\phi_{n-1}, (B - B_a)\phi_{n-1}\rangle$$
$$- \tfrac{1}{2}\langle\psi_n - \psi_{n-1}, (C_a - C)(\psi_n - \psi_{n-1})\rangle \qquad (V.4.7)$$

where $A\psi_n + B_a\phi_n + f = (B_a - B)\phi_{n-1}$ (V.4.8)

and $A^x\phi_n - C_a\psi_n + g = (C - C_a)\psi_{n-1}$ (V.4.9)

We want to find conditions on the operators in the last two equations which will guarantee convergence of the iterative scheme to (ϕ_e, ψ_e), where

$A\psi_e + B\phi_e + f = 0$ (V.4.10)

and $A^x\phi_e - C\psi_e + g = 0$ (V.4.11)

As in the sections dealing with convergence in the last chapter, sections IV.4 and IV.5, we shall use theorem (II.16.1).

By this theorem, we need to express the iteration equations (V.4.8) and (V.4.9) either in the form $\phi_n - \phi_e = P(\phi_{n-1} - \phi_e)$ and $\psi_n - \psi_e = R(\phi_n - \phi_e)$, or in the form $\psi_n - \psi_e = P(\psi_{n-1} - \psi_e)$ and $\phi_n - \phi_e = R(\psi_{n-1} - \psi_e)$; then $\lim_{n\to\infty} \{ \|\phi_n - \phi_e\| + \|\psi_n - \psi_e\| \} = 0$ if P is a linear, self-adjoint operator and there exist real numbers q and Q such that

$$-I < qI \leq P \leq QI < I \qquad (V.4.12)$$

If convergence to (ϕ_e, ψ_e) can be proved, then it also follows that

$$\lim_{n\to\infty} J(\phi_n, \psi_n) = L(\phi_e, \psi_e).$$

We shall assume that B, C, Ba and Ca are bounded below and therefore have inverses B^{-1}, C^{-1}, Ba^{-1} and Ca^{-1} respectively. In order to express equations (V.4.8) and (V.4.9) in the form $\phi_n - \phi_e = P(\phi_{n-1} - \phi_e)$ or $\psi_n - \psi_e = P(\psi_{n-1} - \psi_e)$, we have to let either C = Ca or B = Ba; these conditions will be dealt with in turn. We then require B $\neq$ Ba or C $\neq$ Ca, respectively.

(a) Let C = Ca; then equations (V.4.8) and (V.4.9) become

$$A\psi_n + Ba\,\phi_n + f = (Ba - B)\phi_{n-1}, \qquad \text{(V.4.13)}$$

and $A^x\phi_n - C\psi_n + g = 0$ (V.4.14)

Substituting equation (V.4.14) into (V.4.13) results in

$$(AC^{-1}A^x + Ba)\phi_n + AC^{-1}g + f = (Ba - B)\phi_{n-1}$$

or $(AC^{-1}A^x + Ba)\phi_n - (AC^{-1}A^x + B)\phi_e = (Ba - B)\phi_{n-1}$,

as $(AC^{-1}A^x + B)\phi_e + AC^{-1}g + f = 0$.

Then we have $(AC^{-1}A^x + Ba)(\phi_n - \phi_e) + (Ba - B)\phi_e = (Ba - B)\phi_{n-1}$

or $$\phi_n - \phi_e = (AC^{-1}A^x + Ba)^{-1}(Ba - B)(\phi_{n-1} - \phi_e) \qquad \text{(V.4.15)}$$

Also from equations (V.4.10) and (V.4.11),

$\psi_n - \psi_e = C^{-1}A^x(\phi_n - \phi_e)$; and hence, by theorem (II.16.1), the iterative scheme specified by equations (V.4.13) and (V.4.14) will converge to (ϕ_e, ψ_e) if the operator $(AC^{-1}A^x + Ba)^{-1}$ exists, if $(AC^{-1}A^x + Ba)^{-1}(Ba - B)$ is linear and self-adjoint, and if there exist q and Q such that

$$-I < qI \leq (AC^{-1}A^x + Ba)^{-1}(Ba - B) \leq QI < I \qquad \text{(V.4.16)}$$

(b) Let B = Ba; then equations (V.4.5) and (V.4.9) become

$$A\psi_n + B\phi_n + f = 0 \qquad \text{(V.4.17)}$$

and $A^x\phi_n - Ca\,\psi_n + g = (C - Ca)\psi_{n-1}$ (V.4.18)

Substituting equation (V.4.17) into (V.4.18) gives

$$-(A^x B^{-1} A + Ca)\psi_n + g - A^x B^{-1} f = (C - Ca)\psi_{n-1}$$

or, as $-(A^x B^{-1} A + C)\psi_e + g - A^x B^{-1} f = 0$,

$$-(A^x B^{-1} A + Ca)\psi_n + (A^x B^{-1} A + C)\psi_e = (C - Ca)\psi_{n-1}$$

Then we have $-(A^x B^{-1} A + Ca)(\psi_n - \psi_e) = (C - Ca)(\psi_{n-1} - \psi_e)$

or $(\psi_n - \psi_e) = -(A^x B^{-1} A + Ca)^{-1}(C - Ca)(\psi_{n-1} - \psi_e)$ (V.4.19)

From equations (V.4.10) and (V.4.11), $\phi_n - \phi_e = -B^{-1} A (\phi_n - \phi_e)$;

therefore by theorem (II.16.1), the iterative scheme specified by equations (V.4.17) and (V.4.18) will converge to (ϕ_e, ψ_e) if the operator $(A^x B^{-1} A + Ca)^{-1}$ exists, if $(A^x B^{-1} A + Ca)^{-1}(C - Ca)$ is linear and self-adjoint and if there exist q and Q such that

$$-I < qI \leq (A^x B^{-1} A + Ca)^{-1}(C - Ca) \leq QI < I \qquad \text{(V.4.20)}$$

<u>Theorem (V.2.4)</u>

From the last section, we define $L_b(\phi, \psi)$ by the equation

$$L_b(\phi, \psi) = \langle \phi, A\psi \rangle + \tfrac{1}{2}\langle \phi, B_b\phi \rangle - \tfrac{1}{2}\langle \psi, C_b\psi \rangle$$
$$+ \langle \phi, f \rangle + \langle \psi, g \rangle \qquad \text{(V.4.21)}$$

where B_b and C_b are symmetric, positive-definite operators such that

$$0 < B \leq B_b \quad \text{and} \quad 0 < C_b \leq C \qquad \text{(V.4.22)}$$

The comparison upper bound is given by

$$K(\phi_1, \psi_1) = \tfrac{1}{2}\langle \phi_1, B\phi_1 \rangle + \tfrac{1}{2}\langle \psi_1, C\psi_1 \rangle + \langle \phi_1, f \rangle$$
$$+ \tfrac{1}{2}\langle \phi_o - \phi_1, (B_b - B)(\phi_o - \phi_1) \rangle$$
$$+ \tfrac{1}{2}\langle \psi_o, (C - C_b)\psi_o \rangle - \tfrac{1}{2}\langle \psi_1, (C - C_b)\psi_1 \rangle \qquad \text{(V.4.23)}$$

where $A\psi_1 + B_b\phi_1 + f = (B_b - B)\phi_o$ (V.4.24)

and $A^x\phi_1 - C_b\psi_1 + g = (C - C_b)\psi_o$ (V.4.25)

We set up a cobweb iterative scheme by letting $(\phi_o, \psi_o) = (\phi_{n-1}, \psi_{n-1})$ and $(\phi_1, \psi_1) = (\phi_n, \psi_n)$; this results in the pair of equations

$$A\psi_n + B_b\phi_n + f = (B_b - B)\phi_{n-1}, \qquad \text{(V.4.26)}$$

$$A^x\phi_n - C_b\psi_n + g = (C - C_b)\psi_{n-1} \qquad \text{(V.4.27)}$$

Comparing this iterative scheme with that for the lower comparison bound, equations (V.4.8) and (V.4.9), it is obvious that the only difference between the two schemes is that the comparison operators B_a and C_a in the lower bound are replaced by B_b and C_b respectively in the upper bound; we can therefore immediately write down the conditions for convergence for the iterative scheme given by equations (V.4.26) and (V.4.27).

The cobweb iterative scheme specified by equations (V.4.26) and (V.4.27) will converge to (ϕ_e, ψ_e) if the following conditions are satisfied:

(a) $C = C_b$, $(AC^{-1} A^x + B_b)^{-1}$ exists, $(AC^{-1} A^x + B_b)^{-1} (B_b - B)$ is linear and self-adjoint, and there exist q and Q such that

$$- I < qI \leq (AC^{-1} A^x + B_b)^{-1} (B_b - B) \leq QI < I \qquad \text{(V.4.28)}$$

or

(b) $B = B_b$, $(A^x B^{-1} A + C_b)^{-1}$ exists, $(A^x B^{-1} A + C_b)^{-1} (C - C_b)$ is linear and self-adjoint, and there exist q and Q such that

$$- I < qI \leq (A^x B^{-1} A + C_b)^{-1} (C - C_b) \leq QI < I \qquad \text{(V.4.29)}$$

Table (V.4.1)

Summary of Conditions for Convergence

Theorem	(V.2.3)	(V.2.4)
Scheme	$A\psi_n + B_a\phi_n + f = (B_a - B)\phi_{n-1}$ $A^x\phi_n - C_a\psi_n + g = (C - C_a)\psi_{n-1}$	$A\psi_n + B_b\phi_n + f = (B_b - B)\phi_{n-1}$ $A^x\phi_n - C_b\psi_n + g = (C - C_b)\psi_{n-1}$
Converges to (ϕ_e, ψ_e) if P is linear, self-adjoint and for q, $Q \in R$, $-I < qI \leq P \leq QI < I$	a) $C = C_a$, $B > B_a > 0$ $P = (AC^{-1}A^x + B_a)^{-1}(B_a - B)$ b) $B = B_a$, $C_a > C > 0$ $P = (A^x B^{-1} A + C_a)^{-1}(C - C_a)$	a) $C = C_b$, $B_b > B > C$ $P = (AC^{-1}A^x + B_b)^{-1}(B_b - B)$ b) $B = B_b$, $C > C_b > 0$ $P = (A^x B^{-1} A + C_b)^{-1}(C - C_b)$

CHAPTER V

We shall consider the convergence conditions in Table (V.4.1) in turn and show briefly that there are operators for which the convergence conditions can be satisfied. In each case we are going to use the integral operator K from the first example in section II.15, which is defined by

$K\phi(x) = \int_0^1 k(x,y)\phi(y)\,dy$, where $k(x,y) = \begin{matrix} y, x \geq y \\ x, x \leq y \end{matrix}$, $x,y \in [0,1]$. K is

a linear, self-adjoint positive-definite operator which satisfies

$$0 < K \leq \frac{4}{\pi^2} I \tag{V.4.30}$$

This also implies that $aK + bI$ is also linear and self-adjoint, where a and b are real numbers, a non-zero. From the above table, the cobweb iterative schemes converge to (ϕ_e, ψ_e) if P is linear and self-adjoint and satisfies

$$- I < qI \leq P \leq QI < I \tag{V.4.31}$$

Theorem (V.2.3)

a) $P = (AC^{-1}A^x + Ba)^{-1}(Ba - B)$, $B > Ba$

Let $A = A^x = C = I$, $B = K + I$ and $Ba = m\,I$ (V.4.32)

Then $I < B < (1 + \frac{4}{\pi^2})I$ and $B > Ba$ if $m \in]0,1[$

$P = \frac{(m-1)I - K}{(1+m)}$: Using equation (V.4.30),

$$\frac{(m - 1 - \frac{4}{\pi^2})I}{m+1} \leq P < \frac{(m-1)}{(m+1)} < I$$

and hence the right hand side of (V.4.31) is satisfied.

The left hand side is satisfied if

$$\frac{(m-1) - \frac{4}{\pi^2}}{m+1} > -1 \quad \text{or} \quad m > \frac{2}{\pi^2}$$

Hence the operators given in equation (V.4.32) satisfy the convergence conditions if $\frac{2}{\pi^2} < m < 1$.

The problem obtained from this choice of operators is

$(K + 2I)\phi_e = -(f + g)$.

CHAPTER V

We shall consider the convergence conditions in Table (V.4.1) in turn and show briefly that there are operators for which the convergence conditions can be satisfied. In each case we are going to use the integral operator K from the first example in section II.15, which is defined by

$K\phi(x) = \int_0^1 k(x,y)\phi(y)\,dy$, where $k(x,y) = \begin{matrix} y, x \geq y \\ x, x \leq y \end{matrix}$, $x,y \in [0,1]$. K is

a linear, self-adjoint positive-definite operator which satisfies

$$0 < K \leq \frac{4}{\pi^2} I \qquad \text{(V.4.30)}$$

This also implies that $aK + bI$ is also linear and self-adjoint, where a and b are real numbers, a non-zero. From the above table, the cobweb iterative schemes converge to (ϕ_e, ψ_e) if P is linear and self-adjoint and satisfies

$$- I < qI \leq P \leq QI < I \qquad \text{(V.4.31)}$$

Theorem (V.2.3)

a) $P = (AC^{-1} A^x + Ba)^{-1} (Ba - B)$, $\quad B > Ba$

Let $A = A^x = C = I$, $\quad B = K + I$ and $Ba = m\,I$ (V.4.32)

Then $I < B < (1 + \frac{4}{\pi^2})I$ and $B > Ba$ if $m \in]0,1[$

$P = \frac{(m - 1)I - K}{(1 + m)}$: Using equation (V.4.30),

$$\frac{(m - 1 - \frac{4}{\pi^2})I}{m + 1} \leq P < \frac{(m - 1)}{(m + 1)} < I$$

and hence the right hand side of (V.4.31) is satisfied.

The left hand side is satisfied if

$$\frac{(m - 1) - \frac{4}{\pi^2}}{m + 1} > -1 \quad \text{or} \quad m > \frac{2}{\pi^2}$$

Hence the operators given in equation (V.4.32) satisfy the convergence conditions if $\frac{2}{\pi^2} < m < 1$.

The problem obtained from this choice of operators is

$(K + 2I)\phi_e = -(f + g)$.

b) $P = (A^x B^{-1} A + Ca)^{-1} (C - Ca), \quad Ca > C$

Let $A = A^x = B = I, \quad C = K$ and $Ca = n\,I$ (V.4.33)

If $n > \frac{4}{\pi^2}$ then as $\frac{4}{\pi^2} I \geqslant K$, $nI > K$ or $Ca > C$ as required.

$P = \frac{K - nI}{(n + 1)}$ so using equation (V.4.30),

$$- I < -\left(\frac{n}{n + 1}\right) I < P \leq \frac{\left(\frac{4}{\pi^2} - n\right) I}{n + 1}$$

and hence the left hand side of equation (V.4.31) is satisfied. The right hand side is satisfied if

$$\frac{\frac{4}{\pi^2} - n}{n + 1} < 1 \quad \text{or} \quad n > \frac{2}{\pi^2} - \frac{1}{2}, \quad \text{As } n > \frac{4}{\pi^2} > \frac{2}{\pi^2} - \frac{1}{2},$$

the right hand side is satisfied and hence the operators given in equation (V.4.33) satisfy the convergence conditions if $n > \frac{4}{\pi^2}$. The problem obtained from this choice of operators is $(I + K)\psi_e = g - f$.

Theorem (V.2.4)

a) $P = (AC^{-1} A^x + Bb)^{-1} (Bb - B), \quad Bb > B$

Let $A = A^x = C = I, \quad B = K$ and $Bb = m\,I$ (V.4.34)

If $m > \frac{4}{\pi^2}$ then as $\frac{4}{\pi^2} I \geqslant K$, $m\,I > K$ or $Bb > B$ as required.

$P = \frac{mI - K}{(m + 1)}$ so using equation (V.4.30),

$$\frac{\left(m - \frac{4}{\pi^2}\right) I}{m + 1} \leq P < \frac{m}{m + 1} I < I$$

and hence the right hand side of equation (V.4.31) is satisfied. The left hand side is satisfied if

$$\frac{m - \frac{4}{\pi^2}}{m + 1} > -1 \quad \text{or} \quad m > \frac{2}{\pi^2} - \frac{1}{2}. \quad \text{As } m > \frac{4}{\pi^2} > \frac{2}{\pi^2} - \frac{1}{2}$$

the left hand side is satisfied and hence the operators given in equation (V.4.34) satisfy the convergence conditions if $n > \frac{4}{\pi^2}$.

The problem obtained from this choice of operators is $(I + K)\phi_e = -(f + g)$.

b) $P = (A^x B^{-1} A + C_b)^{-1} (C - C_b), \quad C > C_b$

Let $A = A^x = B = I, \quad C = K + I$ and $C_b = n I$ (V.4.35)

Then $I < C \leqslant (1 + \frac{4}{\pi^2})I$ and $C > C_b$ if $n \in \,]0,1[$.

$P = \frac{K - (n - 1)}{(1 + n)} I$; using equation (V.4.30),

$$- I < \frac{(1 - n)}{1 + n} I < P < \frac{(\frac{4}{\pi^2} - (n - 1) I}{n + 1}$$

and hence the left hand side of equation (V.4.31) is satisfied. The right hand side is satisfied if $\frac{\frac{4}{\pi^2} - (n - 1)}{n + 1} < 1$ or $n > \frac{2}{\pi^2}$

hence the operators given in equation (V.4.35) satisfy the convergence conditions if $\frac{2}{\pi^2} < n < 1$.

The problem obtained for this choice of operators is $(K + 2I)\psi_e = g - f$.

Of course, we can still apply the iterative schemes to problems for which we cannot prove convergence, and it may not be necessary to take any comparison operators equal to operators in $L(\phi , \psi)$, this will be briefly considered here.

<u>Theorem (V.2.3)</u>

From equations (V.3.16) to (V.3.18), we can set up an iterative scheme by letting $(\phi_o, \psi_o) = (\phi_{n-1}, \psi_{n-1})$ and $(\phi_1, \psi_1) = (\phi_n, \psi_n)$. We then have

$L(\phi_e, \psi_e) \geqslant J(\phi_n, \psi_n)$

where $J(\phi_n, \psi_n) = -\frac{1}{2}\langle \phi_n, B\phi_n\rangle - \frac{1}{2}\langle \psi_n, C\psi_n\rangle + \langle \psi_n, g\rangle$

$$+ \tfrac{1}{2}\langle \phi_n, (B - Ba)\phi_n\rangle - \tfrac{1}{2}\langle \phi_{n-1}, (B - Ba)\phi_{n-1}\rangle$$

$$- \tfrac{1}{2}\langle \psi_{n-1} - \psi_n, (Ca - C)(\psi_{n-1} - \psi_n)\rangle \qquad (V.4.36)$$

where $A\psi_n + Ba\phi_n + f = (Ba - B)\phi_{n-1}$ (V.4.37)

$A^x\phi_n - Ca\psi_n + g = (C - Ca)\psi_{n-1}$ (V.4.38)

Theorem (V.2.4)

From equations (V.3.23) to (V.3.25) we can set up an iterative scheme by letting $(\phi_o, \psi_o) = (\phi_{n-1}, \psi_{n-1})$ and $(\phi_1, \psi_1) = (\phi_n, \psi_n)$. We then have

$L(\phi_e, \psi_e) \leq K(\phi_n, \psi_n)$

$$\text{where } K(\phi_n, \psi_n) = \tfrac{1}{2}\langle \phi_n, B\phi_n \rangle + \tfrac{1}{2}\langle \psi_n, C\psi_n \rangle + \langle \phi_n, f \rangle + \tfrac{1}{2}\langle \phi_{n-1} - \phi_n, (B_b - B)(\phi_{n-1} - \phi_n) \rangle + \tfrac{1}{2}\langle \psi_{n-1}, (C - C_b)\psi_{n-1} \rangle - \tfrac{1}{2}\langle \psi_n, (C - C_b)\psi_n \rangle \tag{V.4.39}$$

$$\text{where } A\psi_n + B_b\phi_n + f = (B_b - B)\phi_{n-1} \tag{V.4.40}$$

$$A^x\phi_n - C_b\psi_n + g = (C - C_b)\psi_{n-1} \tag{V.4.41}$$

It has not been possible to show for the two iterations given above, that either $\lim_{n\to\infty} \{\|\phi_n - \phi_e\| + \|\psi_n - \psi_e\|\} = 0$, or $\lim_{n\to\infty} J(\phi_n, \psi_n) = L(\phi_e, \psi_e)$ and $\lim_{n\to\infty} K(\phi_n, \psi_n) = L(\phi_e, \psi_e)$, or $\lim_{n\to\infty} \{K(\phi_n, \psi_n) - J(\phi_n, \psi_n)\} = 0$; this does not mean, however, that there are no problems for which one or more of these convergence conditions is satisfied.

CHAPTER V

V.5 Applications

In this section we shall apply the methods developed in the previous sections to two particular examples.

In the first example we apply a cobweb iterative scheme to an integral equation problem using theorem (V.2.4). We show that convergence can be proved for a cobweb iterative scheme involving a comparison operator, but we cannot prove convergence for cobweb iterative schemes without involving the use of a comparison operator. Finally, the iterations are carried out and shown to converge to the exact solution of the equation.

For the second example we use the same integral equation problem as in example 1. This time we shall show that the conditions given at the end of section V.3, for the comparison bound given in theorem (V.2.3) to be better than the classical bound, can be satisfied; we shall also demonstrate that a reasonable bound can be obtained.

CHAPTER V

Example 1

Consider the problem

$$(I + K)\,\phi_e(x) = \tfrac{1}{2} \qquad \text{(V.5.1)}$$

where K is the integral operator specified by the equations

$$K\phi(x) = \int_0^1 k(x,y)\phi(y)\, dy$$

$$\text{with } k(x,y) = \begin{matrix} y, & x \geq y \\ x, & x \leq y \end{matrix} \qquad x,y \in \,]0,1[\qquad \text{(V.5.2)}$$

and I is the identity operator.

From section II.15, K is a linear, self-adjoint operator with bounds

$$0 < K \leq \frac{4}{\pi^2} I \qquad \text{(V.5.3)}$$

Equation (V.5.1) arises from the gradients of the quadratic functional given in equation (V.3.1), if we specify the operators and functions f and g as

$$A = A^x = \frac{1}{\sqrt{2}} I, \quad B = I, \quad C = \tfrac{1}{2}I + K, \quad f = 0 \text{ and } g = -\frac{\sqrt{2}}{2} \qquad \text{(V.5.4)}$$

Now, from table (IV.4.1), the two possible cobweb iterative schemes involving the specification in equation (V.5.4) are

A) $\frac{1}{\sqrt{2}}\psi_n + \phi_n = 0, \quad \frac{1}{\sqrt{2}}\phi_{n+1} - (\tfrac{1}{2}I + K)\psi_n - \frac{\sqrt{2}}{2} = 0$

and

B) $\frac{1}{\sqrt{2}}\psi_n + \phi_n = 0, \quad \frac{1}{\sqrt{2}}\phi_n - (\tfrac{1}{2}I + K)\psi_{n+1} - \frac{\sqrt{2}}{2} = 0.$

Iteration A will converge to (ϕ_e, ψ_e) if there exists q and Q such that

$$-I < qI \leq -\left(\frac{1}{\sqrt{2}}\right)^{-1}(\tfrac{1}{2}I + K)\left(\frac{1}{\sqrt{2}}\right)^{-1} \leq QI < I \qquad \text{(V.5.5)}$$

and iteration B will converge to (ϕ_e, ψ_e) if there exists q and Q such that

$$-I < qI \leq -\left(\frac{1}{\sqrt{2}}\right)(\tfrac{1}{2}I + K)^{-1}\left(\frac{1}{\sqrt{2}}\right) \leq QI < I \qquad \text{(V.5.6)}$$

Iteration B can be eliminated immediately: the inverse of the operator $(\tfrac{1}{2}I + K)$ is a differential operator, which is not bounded, and hence we cannot find q and Q to satisfy equation (V.5.6).

Equation (V.5.5) can be rewritten as

$$- I < - QI \leq 2(\tfrac{1}{2}I + K) \leq - qI < I \qquad (V.5.7)$$

From equation (V.5.3),

$$I < 2(\tfrac{1}{2}I + K) \leq (1 + \frac{2}{\pi^2}) I$$

and therefore we cannot satisfy equation (V.5.7), and so cannot prove convergence for iteration A.

We now turn to the iterative schemes obtained in the last section for the quadratic functional using theorem (V.2.4). If we let $B = B_b$ then, from equations (V.4.26) and (V.4.27), the iteration with the specification given in equation (V.5.4) is given by the pair of equations

$$\frac{1}{\sqrt{2}} \psi_n + \phi_n = 0, \qquad (V.5.8)$$

$$\frac{1}{\sqrt{2}} \phi_n - C_b \psi_n - \frac{\sqrt{2}}{2} = (\tfrac{1}{2}I + K - C_b)\psi_{n-1} \qquad (V.5.9)$$

where C_b is a linear, symmetric operator satisfying

$$(\tfrac{1}{2}I + K) > C_b > 0. \qquad (V.5.10)$$

The choice $C_b = \frac{1}{2}I$ satisfies the last equation; then for convergence of the iterative scheme given in equations (V.5.8) and (V.5.9), we require there to exist real numbers q and Q such that

$$- I < qI \leq \left\{\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right) I + \frac{1}{2} I\right\}^{-1} \left\{\frac{1}{2} I + K - \frac{1}{2} I\right\} \leq QI < I \qquad (V.5.11)$$

Equation (V.5.11) can be rewritten as

$$- I < qI \leq K \leq QI < I \qquad (V.5.12)$$

and, as $0 < K \leq \frac{4}{\pi^2} I < I$, equation (V.5.12) is satisfied and convergence of the iterative scheme will occur.

With $C_b = \frac{1}{2} I$, the iterative scheme can be written as

$$\frac{1}{\sqrt{2}} \psi_n + \phi_n = 0, \quad \frac{1}{\sqrt{2}}\phi_n - \tfrac{1}{2}\psi_n - \frac{\sqrt{2}}{2} = K\psi_{n-1}, \quad \text{or}$$

$$\frac{1}{\sqrt{2}} \phi_n + \frac{\sqrt{2}}{2} \phi_n - \frac{\sqrt{2}}{2} = K (- \sqrt{2}\phi_{n-1})$$

that is $\phi_n = - K\phi_{n-1} + \frac{1}{2}$ (V.5.13)

CHAPTER V

We can start with any arbitrary initial function ϕ_o; we will take the simplest, $\phi_o = 0$. Then the iterative sequence is:

$$\phi_o = 0$$

$$\phi_1 = - K(0) + \tfrac{1}{2} = \tfrac{1}{2}$$

$$\phi_2 = - K(\tfrac{1}{2}) + \tfrac{1}{2} = \tfrac{1}{2}(1 - K(1))$$

$$\phi_3 = - K \ \tfrac{1}{2}(1 - K(1)) + \tfrac{1}{2} = \tfrac{1}{2}(1 - K(1) + K^2(1))$$

$$\phi_4 = - K \ \tfrac{1}{2}(1 - K(1) + K^2(1)) + \tfrac{1}{2}$$

$$= \tfrac{1}{2}(1 - K(1) + K^2(1) - K^3(1))$$

$$\phi_{n+1} = \tfrac{1}{2}(1 - K(1) + K^2(1) - K^3(1) + \dots + (-1)^n \ K^n(1)) \qquad (V.5.14)$$

where $K^2(1)$ means $K(K(1))$, $K^3(1) = K(K^2(1))$, and so on.

Using equation (V.5.2), and example 1 of section II.15,

$$K(x^m) = \int_o^x y\, y^m\, dy + \int_x^1 x\, y^m\, dy$$

$$= \frac{x}{m+1} - \frac{x^{m+2}}{(m+1)(m+2)} \qquad (V.5.15)$$

Using equation (V.5.15),

$$K(1) = a_2 x - \frac{x^2}{2!}$$

$$K^2(1) = a_4 x - \frac{a_2 x^3}{3!} + \frac{x^4}{4!}$$

$$K^3(1) = a_6 x - \frac{a_4 x^3}{3!} + \frac{a_2 x^5}{5!} - \frac{x^6}{6!}$$

$$K^4(1) = a_8 x - \frac{a_6 x^3}{3!} + \frac{a_4 x^5}{5!} - \frac{a_2 x^7}{7!} + \frac{x^8}{8!}$$

$$K^n(1) = a_{2n} x - \frac{a_{2n-2} x^3}{3!} + \frac{a_{2n-4} x^5}{5!} - \frac{a_{2n-6} x^7}{7!} + \dots$$

$$\dots + \frac{(-1)^{n-1} a_2 x^{2n-1}}{(2n-1)!} + \frac{(-1)^n x^{2n}}{(2n)!}$$

(V.5.16)

where $a_2 = 1$

$$a_4 = \frac{a_2}{2!} - \frac{1}{3!}$$

$$a_6 = \frac{a_4}{2!} - \frac{a_2}{4!} + \frac{1}{5!}$$

$$a_8 = \frac{a_6}{2!} - \frac{a_4}{4!} + \frac{a_2}{6!} - \frac{1}{7!}$$

$$a_{2n} = \frac{a_{2n-2}}{2!} - \frac{a_{2n-4}}{4!} + \frac{a_{2n-6}}{6!} + \cdots + \frac{(-1)^{n-1}\, a_4}{(2n-4)!} + \frac{(-1)^n\, a_2}{(2n-2)!} + \frac{(-1)^{n-1}}{(2n-1)!} \qquad \text{(V.5.17)}$$

Using these terms for $K(1)$, $K^2(1)$, ... $K^n(1)$ gives

$$\begin{aligned}\phi_{n+1} = \tfrac{1}{2}\Big\{ 1 &- a_2 x + \frac{x^2}{2!} \\ &+ a_4 x - \frac{a_2}{3!}x^3 + \frac{x^4}{4!} \\ &- a_6 x + \frac{a_4}{3!}x^3 - \frac{a_2}{5!}x^5 + \frac{x^6}{6!} \\ &+ a_8 x - \frac{a_6}{3!}x^3 + \frac{a_4}{5!}x^5 - \frac{a_2}{7!}x^7 \\ &+ \frac{x^8}{8!} \\ &+ \ldots\ldots \Big\} \quad \text{or}\end{aligned}$$

$$\begin{aligned}\lim_{n\to\infty} \phi_{n+1} &= \tfrac{1}{2}\sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \\ &- \tfrac{1}{2}(a_2 - a_4 + a_6 - a_8 \ldots)\, x \\ &- \tfrac{1}{2}(a_2 - a_4 + a_6 - a_8 \ldots)\, \frac{x^3}{3!} \\ &- \tfrac{1}{2}(a_2 - a_4 + a_6 - a_8 \ldots)\, \frac{x^5}{5!} \quad \ldots\ldots\end{aligned}$$

That is,

$$\begin{aligned}\lim_{n\to\infty} \phi_{n+1} &= \tfrac{1}{2}\sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \\ &- \tfrac{1}{2}(a_2 - a_4 + a_6 - a_8 \ldots) \sum_{n=0}^{\infty} \frac{x^{2n-1}}{(2n-1)!} \qquad \text{(V.5.18)}\end{aligned}$$

Using equation (V.5.17), we have

$a_2 = 1, \quad a_4 = \frac{1}{3}, \quad a_6 = \frac{2}{15}, \quad a_8 = \frac{17}{315};$

it can be shown that the first ten terms of the sequence $a_2 - a_4 + a_6 - a_8 \ldots$ are the first ten terms in the series for tanh 1 ((40) pages 35, 1079).

We suspect, therefore, that

$\tanh 1 = \sum_{n=1}^{\infty} (-1)^{n-1} a_{2n} (1)^{2n-1}$ and we can show that this is true if we show that

$$\tanh x = \sum_{n=1}^{\infty} (-1)^{n-1} a_{2n} x^{2n-1} \qquad \text{(V.5.19)}$$

where a_{2n} is given by equation (V.5.17).

The usual series representation of tanh x is, from page 35 of (40),

$$\tanh x = \sum_{n=1}^{\infty} \frac{2^{2n}(2^{2n}-1)}{(2n)!} B_{2n} x^{2n-1} \qquad \text{(V.5.20)}$$

where B_{2n} are the Bernoulli numbers.

To show that equation (V.5.19) is correct, we need to find b_{2n} such that

$$(-1)^{n-1} a_{2n} = \frac{2^{2n}(2^{2n}-1) b_{2n}}{(2n)!} \qquad \text{(V.5.21)}$$

and then prove $b_{2n} = B_{2n} \; \forall \; n$.

$$b_{2n} = \frac{(-1)^{n-1}(2n)!\, a_{2n}}{2^{2n}(2^{2n}-1)}$$

$$= \frac{(-1)^{n-1}(2n)!}{2^{2n}(2^{2n}-1)} \left(\frac{a_{2n-2}}{2!} - \frac{a_{2n-4}}{4!} + \frac{a_{2n-6}}{6!} + \ldots + \frac{(-1)^{n-1} a_4}{(2n-4)!} + \frac{(-1)^n a_2}{(2n-2)!} + \frac{(-1)^{n-1}}{(2n-1)!} \right) \quad \text{(from equation (V.5.17))}$$

$$= \frac{(-1)^{n-1}(2n)!}{2^{2n}(2n-1)} \left(\frac{2^{2n-2}(2^{2n-2}-1) b_{2n-2}}{(-1)^{n-2}(2n-2)!\, 2!} - \frac{2^{2n-4}(2^{2n-4}-1) b_{2n-4}}{(-1)^{n-3}(2n-4)!\, 4!} \right.$$

$$\left. + \frac{2^{2n-6}(2^{2n-6}-1) b_{2n-6}}{(-1)^{n-1}(2n-6)!\, 6!} + \ldots + \frac{(-1)^{n-1} 2^4 (2^4-1) b_4}{(-1)\, 4!\, (2n-4)!} \right.$$

$$+ \frac{(-1)^n\ 2^2\ (2^2 - 1) b_2}{2\ (2n - 2)!} + \left.\frac{(-1)^{n-1}}{(2n - 1)!}\right\} \quad \text{(using equation (V.5.21))}$$

That is, $$b_{2n} = \frac{2n}{2^{2n}\ (2^{2n} - 1)} - \sum_{m=1}^{n-1} \binom{2n}{2m} \frac{2^{2m}\ (2^{2m} - 1) b_{2m}}{2^{2n}\ (2^{2n} - 1)}$$

$$n = 1, 2, \ldots. \qquad \text{(V.5.22)}$$

From page 804 of (1), the Bernoulli numbers satisfy

$$B_{2n}\ (x + h) = (B\ (x) + h)^{2n}$$

$$= \binom{2n}{0} B_0(x)\ h^{2n} + \binom{2n}{1} B_1(x)\ h^{2n-1} + \binom{2n}{2} B_2(x)\ h^{2n-2}$$

$$+ \ldots\ldots + \binom{2n}{2n} B_{2n}(x)\ h^0 \qquad \text{(V.5.23)}$$

Let $x = 0$ in the last equations; then $B_{2n}(x) = B_{2n}(0) = B_{2n}$.

$$B_{2n}\ (h) = \binom{2n}{0} B_0\ h^{2n} + \binom{2n}{1} B_1\ h^{2n-1} + \binom{2n}{2} B_2\ h^{2n-2}$$

$$+ \ldots\ldots + \binom{2n}{2n} B_{2n}\ h^0 \qquad \text{(V.5.24)}$$

From page 1077 of (40),

$$B_0 = 1, \quad B_1 = -\tfrac{1}{2}, \quad B_{2k+1} = 0,\ k = 1, 2, \ldots. \qquad \text{(V.5.25)}$$

Therefore

$$B_{2n}\ (h) = h^{2n} - n\ h^{2n-1} + B_{2n} + \sum_{m=1}^{n-1} \binom{2n}{2m} h^{2n-2m}\ B_{2m} \qquad \text{(V.5.26)}$$

Let $h = \frac{1}{4}$ in equation (V.5.26), then

$$B_{2n}\ (\tfrac{1}{4}) = \frac{1}{2^{4m}} - \frac{n}{2^{4m-2}} + B_{2n} + \sum_{m=1}^{n-1} \binom{2n}{2m} \frac{1}{2^{4n-4m}} B_{2m}$$

$$\text{or } 2^{4m}\ B_{2n}\ (\tfrac{1}{4}) = 1 - 4n + 2^{4n}\ B_{2n} + \sum_{m=1}^{n-1} \binom{2n}{2m} 2^{4m}\ B_{2m} \qquad \text{(V.5.27)}$$

Let $h = \frac{1}{2}$ in equation (V.5.26), then

$$B_{2n}\ (\tfrac{1}{2}) = \frac{1}{2^{2n}} - \frac{n}{2^{2n-1}} + B_{2n} + \sum_{m=1}^{n-1} \binom{2n}{2m} \frac{1}{2^{2n-2m}} B_{2m}$$

$$\text{or } 2^{2n}\ B_{2n}\ (\tfrac{1}{2}) = 1 - 2n + 2^{2n}\ B_{2n} + \sum_{m=1}^{n-1} \binom{2n}{2m} 2^{2m}\ B_{2m} \qquad \text{(V.5.28)}$$

From equations (V.5.27) and (V.5.28),

$$2^{4n} B_{2n} (\tfrac{1}{4}) - 2^{2n} B_{2n} (\tfrac{1}{2})$$

$$= - 2n + 2^{2n} (2^{2n} - 1) B_{2n} + \sum_{m=1}^{n-1} \binom{2n}{2m} 2^{2m} (2^{2m} - 1) B_{2m}$$

From page 804, (/),

$$B_{2n} (mx) = m^{2n-1} \sum_{k=0}^{m-1} B_{2n} (x + \tfrac{k}{m}) \qquad \text{(V.5.30)}$$

Let m = 4 and x = 0 in this last equation; then

$$B_{2n} = 4^{2n-1} \sum_{k=0}^{3} B_{2n} (\tfrac{k}{4})$$

$$= 2^{4n-2} (B_{2n} + B_{2n} (\tfrac{1}{4}) + B_{2n} (\tfrac{1}{2}) + B_{2n} (\tfrac{3}{4}))$$

$$= 2^{4n-2} (B_{2n} + B_{2n} (\tfrac{1}{2}) + 2 B_{2n} (\tfrac{1}{4})) \qquad \text{(V.5.31)}$$

(using the symmetry relationship, page 804 of (/))

From page 805 of (41),

$$B_{2n} (\tfrac{1}{2}) = - (1 - 2^{1-2n}) B_{2n} \qquad \text{(V.5.32)}$$

Substituting this equation into the previous one,

$$B_{2n} = 2^{4n-2} (B_{2n} - B_{2n} + 2^{1-2n} B_{2n} + 2 B_{2n} (\tfrac{1}{4}))$$

or $$2^{4n} B_{2n} (\tfrac{1}{4}) = (2 - 2^{2n}) B_{2n} \qquad \text{(V.5.33)}$$

also, from (V.4.32), $2^{2n} B_{2n} (\tfrac{1}{2}) = (2 - 2^{2n}) B_{2n}$ (V.5.34)

Hence $2^{4n} B_{2n} (\tfrac{1}{4}) - B^{2n} B_{2n} (\tfrac{1}{2}) = 0$

Therefore, from equation (V.5.29),

$$- 2n + 2^{2n} (2^{2n} - 1) B_{2n} + \sum_{m=1}^{n-1} \binom{2n}{2m} 2^{2m} (2^{2m} - 1) B_{2m} = 0$$

or $$B_{2n} = \frac{2n}{2^{2n} (2^{2n} - 1)} - \sum_{m=1}^{n-1} \binom{2n}{2m} \frac{2^{2m} (2^{2m} - 1) B_{2m}}{2^{2n} (2^{2n} - 1)}$$

$= b_{2n}$ as required

Hence $$\tanh x = \sum_{n=1}^{\infty} (-1)^{n-1} a_{2n} x^{2n-1} \qquad \text{(V.5.35)}$$

where a_{2n} is given by equation (V.5.17)

and so $$\tanh 1 = \sum_{n=1}^{\infty} (-1)^{n-1} a_{2n} \qquad \text{(V.5.36)}$$

Therefore from equation (V.5.18),

$$\lim_{n\to\infty} \phi_{n+1} = \frac{1}{2} \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} - \frac{1}{2} \tanh 1 \sum_{n=0}^{\infty} \frac{x^{2n-1}}{(2n-1)!}$$

$$= \frac{1}{2} \cosh x - \frac{1}{2} \tanh 1 \sinh x \qquad (V.5.37)$$

(From page 35 of (40))

We can check that the iteration has found the solution of the problem given by equations (V.5.1) and (V.5.2) by substituting $\phi_e = \frac{1}{2} \cosh x - \frac{1}{2} \tanh 1 \sinh x$ into these two equations; this check shows that $\lim_{n\to\infty} \phi_{n+1} = \phi_e$, and so the iteration involving the comparison theorem (V.2.4) has converged to the exact solution of the equation, as predicted.

Example 2

We consider again the integral equation problem used in the last example,

$$(I + K)\psi_e(x) = \tfrac{1}{2}, \qquad (V.5.38)$$

where the integral operator K is defined by the equations

$$K\psi(x) = \int_0^1 k(x,y)\,\psi(y)\,dy \qquad (V.5.39)$$

$$\text{where } k(x,y) = \begin{matrix} y, x \geq y \\ x, x \leq y \end{matrix}, \quad x \in \,]0,1[$$

and I is the identity operator.

The inner product is given by

$$\langle h_1(x), h_2(x)\rangle = \int_0^1 h_1(x)\; h_2(x)\; dx \qquad (V.5.40)$$

We are going to apply theorem (V.2.3) to the problem.

Theorem (V.2.3)

In equation (V.3.1), let

$$A = A^x = B = I, \quad C = K, \quad f = 0 \text{ and } g = \tfrac{1}{2}; \qquad (V.5.41)$$

using equations (V.3.11), this results in the problem

$(I + K)\psi_e(x) = \frac{1}{2}$, as required.

From section (V.3), to apply theorem (V.2.3), and to ensure that the comparison bound is better than the classical bound, we require Ba and Ca such that $Ba = B = I$ and $Ca > C$. As $C = K$ satisfies $K \leq \frac{4}{\pi^2}$, we can take Ca as mI, where m is a real number such that $m \geq \frac{4}{\pi^2}$. We then have that the comparison bound $J(\phi_1, \psi_1)$ is greater than the classical bound $L(\phi, \psi)$ if ψ_0 in the comparison bound is taken as equal to ψ_β in the classical bound. Using equations (V.3.26), (V.3.27) and (V.3.36) to (V.3.38), these bounds are given by the equations

$$L(\phi_\beta, \psi_\beta) = -\tfrac{1}{2}\int_0^1 (\phi_\beta{}^2 + \psi_\beta K\psi_\beta - \psi_\beta)\,dx \qquad (V.5.42)$$

$$\text{where } \phi_\beta + \psi_\beta = 0 \qquad (V.5.43)$$

and $J(\phi_1, \psi_1) = -\frac{1}{2}\int_0^1 \{\phi_1^2 + \phi_1 K \psi_1 - \psi_1$

$$+ (\psi_0 - \psi_1)(mI - K)(\psi_0 - \psi_1)\} dx \qquad (V.5.44)$$

where $\psi_1 + \phi_1 = 0$ and $\phi_1 - K\psi_0 + \frac{1}{2} + m(\psi_0 - \psi_1) = 0$ (V.5.45)

These equations can be simplified to

$$L(\phi_s, \psi_s) = -\frac{1}{2}\int_0^1 \{\psi_s(K + I)\psi_s - 1\} dx \qquad (V.5.46)$$

and $J(\phi_1, \psi_1) = -\frac{1}{2}\int_0^1 \{\psi_1(K + I)\psi_1 - 1$

$$+ (\psi_0 - \psi_1)(mI - K)(\psi_0 - \psi_1)\} dx \qquad (V.5.47)$$

where $(1 + m)\psi_1 = (mI - K)\psi_0 + \frac{1}{2}$ (V.5.48)

We choose the simplest trial function, $\psi_s = \psi_0 = 0$ then $L(\phi_s, \psi_s) = 0$

From equation (V.5.45), $\psi_1 = \frac{1}{2(1+m)}$; then $J(\phi_1, \psi_1)$ is given by

$$J(\phi_1, \psi_1) = -\frac{1}{2}\int_0^1 \left\{\left(\frac{1}{2(1+m)}\left(K\left(\frac{1}{2(1+m)}\right) + \frac{1}{2(1+m)}\right) - \frac{1}{2(1+m)}\right.\right.$$

$$\left. + \frac{m}{4(1+m)^2} - \frac{1}{2(1+m)} K\left(\frac{1}{2(1+m)}\right)\right\} dx$$

$$= \frac{1}{8(1+m)} \qquad (V.5.49)$$

Now, as $m \geqslant \frac{4}{\pi^2}$, $\frac{1}{8(1+m)} \leqslant \frac{\pi^2}{8(\pi^2 + 4)}$; and hence we can take

$$J(\phi_1, \psi_1) = \frac{\pi^2}{8(\pi^2 + 4)} \qquad (V.5.50)$$

which is better than the classical lower bound of zero.

From the last example, the exact solution of the equation defined by (V.5.38) and (V.5.39) is

$\psi_e(x) = \frac{1}{2}\cosh x - \frac{1}{2}\tanh 1 \sinh x$; the stationary value $L(\phi_e, \psi_e)$ is then given by

$$L(\phi_e, \psi_e) = \tfrac{1}{2}\langle\phi_e, f\rangle + \tfrac{1}{2}\langle\psi_e, g\rangle$$

$$= \frac{1}{8}\int_0^1 \{\cosh x - \tanh 1 \sinh x\} dx$$

$$= \frac{1}{8}\tanh 1 \qquad (V.5.51)$$

Then the stationary value is approximately equal to 0.095199269, and the lower bound given in equation (V.5.50) is about 0.088949945; the difference is of order 10^{-3}, which is close considering the crudity of the trial function.

We could, of course, use theorem (V.2.4) for this problem, but the choice $A = A^x = C = I$, $B = K$, $f = -\frac{1}{2}$ and $g = 0$, with $B_b = mI \geq \frac{4}{\pi^2} I$ and $C_b = I$, with $\phi_0 = \phi_\alpha = 0$, while satisfying the conditions from section (V.3) for $K(\phi_1, \psi_1)$ to be a better bound than $L(\phi_\alpha, \psi_\alpha)$, does not lead to anything new; in fact what it results in is

$$L(\phi_\alpha, \psi_\alpha) = 0 \qquad K(\phi_1, \psi_1) = -\frac{\pi^2}{8(\pi^2 + 4)} \geq L(\phi_e, \psi_e) = -\tfrac{1}{8} \tanh 1.$$

CHAPTER VI

VI.1 Introduction

In the previous chapters we have been finding bounds to the stationary value $L(\phi_e, \psi_e)$, where we have always taken $L(\phi, \psi)$ as a saddle functional; that is, we have been working with $L(\phi, \psi) : E^1 \times F^1 \to \mathbb{R}$, where E^1 and F^1 are vector spaces and $L(\phi, \psi)$ has been convex in ϕ for all $\psi \in V_1 \subseteq F^1$ and concave in ψ for all $\phi \in U_1 \subseteq E^1$. This implies that ϕ_e, ϕ_α and ϕ_β all belong to the same space U_1 and ψ_e, ψ_α and ψ_β all belong to the same space V_1, where $L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha)$ (VI.1.1)

We intend to show in this final chapter that if instead we assume that $L(\phi, \psi)$ is convex in a set U_2 and concave in a set V_2, where $U_1 \subseteq U_2 \subseteq E^1$ and $V_1 \subseteq V_2 \subseteq F^1$, then we can still obtain the bounds (VI.1.1), but we only need that $\phi_\beta \in U_2$ and $\psi_\alpha \in V_2$, with ϕ_e, $\phi_\alpha \in U_1$ and ψ_e, $\psi_\beta \in V_1$. These convex/concave bounds could be useful; for instance, if U_2 is the set $\{\phi : \phi \in C^1(0,1)\}$ and U_1 is the set $\{\phi : \phi \in C^1(0,1),\ \phi \geq 0\}$, then, unlike the bounds obtained when $L(\phi, \psi)$ is a saddle functional, ϕ_β does not have to be a positive function.

Of course, as $(\phi_e, \psi_e) \in U_1 \times V_1$, a trial function ϕ_β which does not belong to U, is not likely to produce a bound as good as that obtained with a trial function which does belong to U_1; on the other hand, in a particular problem it might be easier to find a trial function ϕ_β which does not belong to U_1, and thus the bounds obtained using the theorem given in this chapter would then be useful.

The chapter consists of two sections, Section VI.2 deals with the basic theory and Section VI.3 looks at three examples.

Examples 1 and 2 are simple examples intending to show that problems to which the theory can be applied do exist; example 1 is concerned with a pair of non-linear simultaneous equations and example 2 looks at a non-linear differential equation.

Example 3 shows that by using theorem (VI.2.1) for a boundary value problem we do not need to include boundary terms in the functional $L(\phi, \psi)$.

VI.2 Dual Extremum Principles for a convex/concave functional

This section starts with the theorem which forms the basis of the chapter, and ends with some comments.

Theorem (VI.2.1)

Let E^1, F^1, U_1, U_2, V_1 and V_2 be vector spaces such that

U_1 is a convex subset of U_2 which is a convex subset of

$E^1 : U_1 \subseteq U_2 \subseteq E^1$ (VI.2.1)

and V_1 is a concave subset of V_2 which is a concave subset of

$F^1 : V_1 \subseteq V_2 \subseteq F^1$ (VI.2.2)

Let $L(\phi, \psi) : E^1 \times F^1 \rightarrow \mathbb{R}$ be a functional which is differentiable at all points (ϕ, ψ) of $E^1 \times F^1$, and is convex in the set U_2 and concave in the set V_2.

Let (ϕ_e, ψ_e) be a stationary value which satisfies the equations

$\nabla_\phi \; L(\phi_e, \psi_e) = 0$ and $\nabla_\psi \; L(\phi_e, \psi_e) = 0$ (VI.2.3)

If there exists a pair (ϕ_β, ψ_β) such that

$\nabla_\phi \; L(\phi_\beta, \psi_\beta) = 0$ (VI.2.4)

and a pair $(\phi_\alpha, \psi_\alpha)$ such that

$\nabla_\psi \; L(\phi_\alpha, \psi_\alpha) = 0$ (VI.2.5)

then $L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha)$ (VI.2.6)

where $\phi_\beta \in U_2$, $\psi_\beta \in V_1$, (VI.2.7)

$\phi_\alpha \in U_1$ and $\psi_\alpha \in V_2$ (VI.2.8)

Proof

By definition (II.8.6), $L(\phi, \psi)$ is convex in the set U_2 if

$L(\phi_1, \psi) - L(\phi_2, \psi) - \langle \phi_1 - \phi_2, \nabla_\phi \; L(\phi_2, \psi) \rangle \geq 0$

$\forall \; (\phi_1, \psi_2) \in U_2$ and $\forall \; \psi \in V_1$ (VI.2.9)

Similarly, by definition (II.8.7), $L(\phi,\psi)$ is concave in the set V_2 if

$$L(\phi,\psi_1) - L(\phi,\psi_2) - \langle \psi_1 - \psi_2, \nabla_\psi L(\phi,\psi_1)\rangle \geq 0$$

$$\forall \phi \in U_1 \text{ and } \forall (\psi_1,\psi_2) \in V_2 \quad \text{(VI.2.10)}$$

Lower Bound

Let $\phi_e \in U_1$ and therefore $\phi_e \in U_2$; (VI.2.11)

Let $\psi_e \in V_1$ and therefore $\psi_e \in V_2$; (VI.2.12)

Let $\phi_s \in U_2$ (VI.2.13)

Let $\psi_s \in V_1$ and therefore $\psi_s \in V_2$; (VI.2.14)

In equation (VI.2.9), let $\phi_1 = \phi_e$, $\phi_2 = \phi_s$ and $\psi = \psi_s$ (VI.2.15)

then $L(\phi_e,\psi_s) - L(\phi_s,\psi_s) - \langle \phi_e - \phi_s, \nabla_\phi L(\phi_s,\psi_s)\rangle \geq 0$

or, as $\nabla_\phi L(\phi_s,\psi_s) = 0$,

$$L(\phi_e,\psi_s) \geq L(\phi_s,\psi_s) \quad \text{(VI.2.16)}$$

In equation (VI.2.10), let $\phi = \phi_e$, $\psi_1 = \psi_e$ and $\psi_2 = \psi_s$; (VI.2.17)

then $L(\phi_e,\psi_e) - L(\phi_e,\psi_s) - \langle \psi_e - \psi_s, \nabla_\psi L(\phi_e,\psi_e)\rangle \geq 0$

or, as $\nabla_\psi L(\phi_e,\psi_e) = 0$,

$$L(\phi_e,\psi_e) \geq L(\phi_e,\psi_s) \quad \text{(VI.2.18)}$$

Putting equations (VI.2.17) and (VI.2.18) together results in

$L(\phi_e,\psi_e) \geq L(\phi_s,\psi_s)$.

Upper Bound

Let $\phi_\alpha \in U_1$ and therefore $\phi_\alpha \in U_2$ (VI.2.19)

Let $\psi_\alpha \in V_2$ (VI.2.20)

In equation (VI.2.9), let $\phi_1 = \phi_\alpha$, $\phi_2 = \phi_e$ and $\psi = \psi_e$ (VI.2.21)

then $L(\phi_\alpha,\psi_e) - L(\phi_e,\psi_e) - \langle \phi_\alpha - \phi_e, \nabla_\phi L(\phi_e,\psi_e)\rangle \geq 0$

or, as $\nabla_\phi L(\psi_e,\psi_e) = 0$

$$L(\phi_e,\psi_e) \leq L(\phi_\alpha,\phi_e) \quad \text{(VI.2.22)}$$

In equation (VI.2.10), let $\phi = \phi_\alpha$, $\psi_1 = \psi_\alpha$ and $\psi_2 = \psi_e$ (VI.2.23)

then $L(\phi_\alpha,\psi_\alpha) - L(\phi_\alpha,\psi_e) - \langle \psi_\alpha - \psi_e, \nabla_\psi L(\phi_\alpha,\psi_\alpha)\rangle \geq 0$

or, as $\nabla_\psi L(\phi_\alpha,\psi_\alpha) = 0$,

$$L(\phi_\alpha,\psi_e) \leq L(\phi_\alpha,\psi_\alpha) \quad \text{(VI.2.24)}$$

Putting equations (VI.2.22) and (VI.2.24) together gives

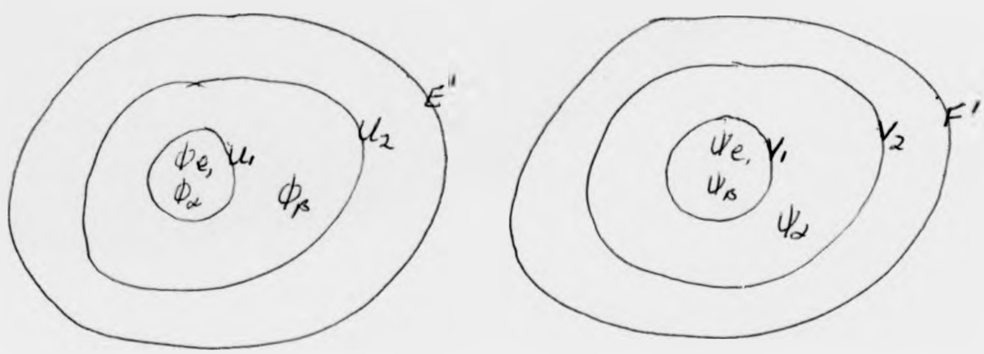
$L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha)$

Hence $L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha)$.

We can illustrate the difference between the bounds given in the usual dual extremum principles and those given in theorem (VI.2.1) by using Venn diagrams; see figures (VI.2.1) and (VI.2.2).

Figure VI.2.1

Figure VI.2.2



As we can see in figure (VI.2.1), in the usual dual extremum principles, ϕ_β, ϕ_α, and ϕ_e all belong to the same convex subset U_1 and ψ_α, ψ_β and ψ_e all belong to the same concave subset V_1, but in the bounds given in theorem (VI.2.1), as we can see in figure (VI.2.2), whereas ϕ_e and ϕ_α belong to the convex subset U_1, ϕ_β need not belong to this set; and similarly ψ_e and ψ_β belong to the concave subset V_1 but ψ_α need not belong to this set. For example, U_2 could be the set of twice differentiable functions and U_1 could be the set of twice differentiable functions which are greater than zero; then ϕ_α and ϕ_e must be twice differentiable functions which are both positive, whereas ϕ_β must be twice differentiable but need not be positive.

It is clear that if $U_1 = U_2$ and $V_1 = V_2$, or alternatively $\phi_\beta \in U$, and $\psi_\alpha \in V$, then the bounds given in theorem (VI.2.1) are the same as the usual dual extremum principles.

If $L(\phi, \psi)$ is strictly convex in the set U_2 and strictly concave in the set V_2, then (ϕ_e, ψ_e) is unique.

CHAPTER VI

VI.3 Examples

Examples 1 and 2 are designed to show that theorem (VI.2.1) does have non-trivial applications.

Example 1

Let $L(\phi,\psi) : U_1 \times V_1 \to \mathbb{R}$ be defined by the equation

$$L(\phi,\psi) = \int_0^1 \left\{ \tfrac{1}{2}\phi^2 - \tfrac{1}{2}\psi^2 + \phi + 14\psi - \phi\psi^2 \right\} dt \qquad (VI.3.1)$$

The functional derivatives are given by

$$\nabla_\phi L(\phi,\psi) = \phi + 1 - \psi^2 \qquad (VI.3.2)$$

$$\nabla_\psi L(\phi,\psi) = -\psi + 14 - 2\phi\psi \qquad (VI.3.3)$$

Using equation (II.9.1), $L(\phi,\psi)$ is a convex/concave saddle functional if

$$\int_0^1 \left\{ \tfrac{1}{2}(\phi_1 - \phi_2)^2 + (\tfrac{1}{2} + \phi_1)(\psi_1 - \psi_2)^2 \right\} dt \geqslant 0 \qquad (VI.3.4)$$

$\forall (\phi_1, \phi_2) \in U_1 \subseteq \mathbb{R}$ and $\forall (\psi_1, \psi_2) \in V_1 \subseteq \mathbb{R}$

It is easy to see from the above equation that we require

$$U_1 = [-\tfrac{1}{2}, \infty[\qquad (VI.3.5)$$

$$\text{and} \quad V_1 = \mathbb{R} \qquad (VI.3.6)$$

Hence the dual extremum principles are (theorem II.12.1)

$L(\phi_\alpha, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha)$, where

$$\phi_e, \phi_\beta, \phi_\alpha \in [-\tfrac{1}{2}, \infty[\qquad (VI.3.7)$$

$$\text{and} \quad \psi_e, \psi_\beta, \psi_\alpha \in \mathbb{R} \qquad (VI.3.8)$$

$$\text{and} \quad \phi_\beta + 1 - \psi_\beta^2 = 0, \quad -\psi_\alpha + 14 - 2\phi_\alpha\psi_\alpha = 0 \qquad (VI.3.9)$$

We now apply theorem (VI.2.1). Using equation (VI.2.9), $L(\phi,\psi)$ is convex in the set U_2 if

$$\int_0^1 \tfrac{1}{2}(\phi_1 - \phi_2)^2 \, dt \geqslant 0 \quad \forall (\phi_1, \phi_2) \in U_2 \qquad (VI.3.10)$$

This gives $U_2 = \mathbb{R}$ (VI.3.11)

Using equation (VI.2.10), $L(\phi, \psi)$ is concave in the set V_2 if

$$\int_0^1 \left(\tfrac{1}{2} + \phi\right) (\psi_1 - \psi_2)^2 \, dt \geq 0 \quad \forall \phi \in U_1 \text{ and}$$

$$\forall \; (\psi_1, \psi_2) \in V_2 \qquad \text{(VI.3.12)}$$

This gives $V_2 = \mathbb{R}$ (VI.3.13)

with U_1 as in equation (VI.3.7).

By theorem (VI.2.1), we then have again $L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\beta)$, but this time

$$\phi_\beta \in \mathbb{R}, \quad \psi_\beta \in \mathbb{R} \qquad \text{(VI.3.14)}$$

$$\phi_\alpha \in \left[-\tfrac{1}{2}, \infty\right[\quad \text{and} \quad \psi_\alpha \in \mathbb{R} \qquad \text{(VI.3.15)}$$

where again $\phi_\beta + 1 - \psi_\beta^2 = 0$ and $-\psi_\alpha + 14 - 2\phi_\alpha \psi_\alpha = 0$.

By comparing equations (VI.3.7) and (VI.3.14), we can see that by using theorem (VI.2.1), ϕ_β is less restricted than it is if theorem (II.12.1) is used.

$L(\phi_\beta, \psi_\beta) = 14\,\psi_\beta - \phi_\beta^2 - \phi_\beta - \frac{1}{2}$

where $\phi_\beta + 1 - \psi_\beta^2 = 0$.

Let $\psi_\beta = m$, m a real number; then $\phi_\beta = m^2 - 1$

and $L(\phi_\beta, \psi_\beta) = -m^4 + m^2 + 14m - \frac{1}{2}$

The maximum of $L(\phi_\beta, \psi_\beta)$ occurs when $m \simeq 1.705$, giving $\phi_\beta = 1.9$, and, in fact the best ϕ_β belongs to $\left[-\frac{1}{2}, \infty\right[$, thus theorem (II.12.1) gives the best result for this problem, as expected.

CHAPTER VI

Example 2

Let $L(\phi, \psi) : E^1 \times F^1 \rightarrow \mathbb{R}$ be defined by the equation

$$L(\phi, \psi) = \int_0^1 \left\{ \tfrac{1}{2}\phi^2 - c\psi^2 - \phi\psi^2 - \phi\frac{d\psi}{dt} + a\phi + b\psi \right\} dt$$

$$+ \phi(1)\psi(1), \quad a, b, c \in \mathbb{R} \qquad \text{(VI.3.16)}$$

where $E^1 = \begin{bmatrix} \phi(t) \\ \phi(1) \\ \phi(0) \end{bmatrix}$ and $F^1 = \begin{bmatrix} \psi(t) \\ \psi(1) \\ \psi(0) \end{bmatrix}$

The gradients are given by the equations

$$\nabla_\phi L(\phi, \psi) = \begin{bmatrix} \phi - \psi^2 - \frac{d\psi}{dt} + a \\ 0 \\ \psi(1) \end{bmatrix} \qquad \text{(VI.3.17)}$$

$$\nabla_\psi L(\phi, \psi) = \begin{bmatrix} -c\psi - 2\phi\psi + \frac{d\phi}{dt} + b \\ \phi(0) \\ 0 \end{bmatrix} \qquad \text{(VI.3.18)}$$

Using equation (II.9.1), $L(\phi, \psi)$ is a convex/concave saddle functional if

$$\int_0^1 \left\{ \tfrac{1}{2}(\phi_1 - \phi_2)^2 + (\phi_1 + \tfrac{1}{2}c)(\psi_1 - \psi_2)^2 \right\} dt \geqslant 0 \qquad \text{(VI.3.19)}$$

$$\forall (\phi_1, \phi_2) \in U_1 \subseteq E^1 \text{ and } \forall (\psi_1, \psi_2) \in V_1 \subseteq F^1$$

Equation (VI.3.19) is satisfied if we take

$$U_1 = \{ \phi : \phi \in C^1(0,1) \text{ and } \phi + \tfrac{1}{2}c \geqslant 0 \} \qquad \text{(VI.3.20)}$$

and $V_1 = \{ \psi : \psi \in C^1(0,1) \}$ (VI.3.21)

Hence the usual dual extremum principles are, using theorem (II.12.1)

$$L(\phi_s, \psi_s) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha)$$

where $\phi_s - \psi_s^2 - \frac{d\psi_s}{dt} + a = 0, \quad \psi_s(1) = 0$ (VI.3.22)

with $\phi_s + \tfrac{1}{2}c \geqslant 0$

and $-c\psi_\alpha - 2\phi_\alpha\psi_\alpha + \frac{d\phi_\alpha}{dt} + b = 0, \quad \phi_\alpha(0) = 0$

with $\phi_\alpha + \frac{1}{2}c \geq 0$ (VI.3.23)

We now apply theorem (VI.2.1). Using equation (VI.2.9), $L(\phi,\psi)$ is convex in the set U_2 if

$$\int_0^1 \tfrac{1}{2}(\phi_1 - \phi_2)^2\, dt \geq 0$$

$$\forall\ (\phi_1, \phi_2) \in U_2$$

Hence we take U_2 as the set

$$U_2 = \{\phi : \phi \in C^1(0,1)\} \tag{VI.3.24}$$

Using equation (VI.2.10), $L(\phi,\psi)$ is concave in the set V_2 if

$$\int_0^1 (\phi + \tfrac{1}{2}c)(\psi_1 - \psi_2)^2\, dt \geq 0$$

$$\forall\ \phi \in U_1 \text{ and } \forall\ (\psi_1, \psi_2) \in V_2 \tag{VI.3.25}$$

We therefore take V_2 as the set

$$V_2 = \{\psi : \psi \in C^1(0,1)\} \tag{VI.3.26}$$

By theorem (VI.2.1), we then have

$$L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha)$$

where $\phi_\beta \in U_2$, $\psi_\beta \in V_1$,

$\phi_\alpha \in U_1$, $\psi_\alpha \in V_2$

That is, we require (ϕ_β, ψ_β) such that

$$\phi_\beta - \psi_\beta{}^2 - \frac{d\psi_\beta}{dt} + a = 0, \quad \psi_\beta(1) = 0 \tag{VI.3.27}$$

and we require $(\phi_\alpha, \psi_\alpha)$ such that

$$-c\psi_\alpha - 2\phi_\alpha\psi_\alpha + \frac{d\phi_\alpha}{dt} + b = 0, \quad \phi_\alpha(0) = 0$$

and $\phi_\alpha + \frac{1}{2}c \geq 0$ (VI.3.28) and $-c\psi_\alpha - 2\phi_\alpha\psi_\alpha + \frac{d\phi_\alpha}{dt} + b = 0, \quad \phi_\alpha(0) = 0$ (VI.3.23)

with $\phi_\alpha + \frac{1}{2}c \geqslant 0$

We now apply theorem (VI.2.1). Using equation (VI.2.9), $L(\phi,\psi)$ is convex in the set U_2 if

$$\int_0^1 \tfrac{1}{2}(\phi_1 - \phi_2)^2\,dt \geqslant 0$$

$$\forall\ (\phi_1,\phi_2) \in U_2$$

Hence we take U_2 as the set

$$U_2 = \{\phi : \phi \in C^1(0,1)\} \tag{VI.3.24}$$

Using equation (VI.2.10), $L(\phi,\psi)$ is concave in the set V_2 if

$$\int_0^1 (\phi + \tfrac{1}{2}c)(\psi_1 - \psi_2)^2\,dt \geqslant 0 \tag{VI.3.25}$$

$$\forall\ \phi \in U_1 \text{ and } \forall\ (\psi_1,\psi_2) \in V_2$$

We therefore take V_2 as the set

$$V_2 = \{\psi : \psi \in C^1(0,1)\} \tag{VI.3.26}$$

By theorem (VI.2.1), we then have

$$L(\phi_\beta,\psi_\beta) \leq L(\phi_e,\psi_e) \leq L(\phi_\alpha,\psi_\alpha)$$

where $\phi_\beta \in U_2$, $\psi_\beta \in V_1$,

$\phi_\alpha \in U_1$, $\psi_\alpha \in V_2$

That is, we require (ϕ_β,ψ_β) such that

$$\phi_\beta - \psi_\beta^2 - \frac{d\psi_\beta}{dt} + a = 0, \quad \psi_\beta(1) = 0 \tag{VI.3.27}$$

and we require $(\phi_\alpha,\psi_\alpha)$ such that

$$-c\psi_\alpha - 2\phi_\alpha\psi_\alpha + \frac{d\phi_\alpha}{dt} + b = 0, \quad \phi_\alpha(0) = 0 \tag{VI.3.28}$$

and $\phi_\alpha + \frac{1}{2}c \geqslant 0$

Comparing equations (VI.3.22) and (VI.3.23), it is easy to see that by using theorem (VI.2.1) we have relaxed the requirement that $\phi_\beta + \frac{1}{2}c \geq 0$.

Let a = - 1, b = 0, c = 1; then

$$L(\phi_\beta, \psi_\beta) = \int_0^1 \left\{ -\tfrac{1}{2}\phi_\beta{}^2 - \tfrac{1}{2}\psi_\beta{}^2 \right\} dt$$

where $\phi_\beta - \psi_\beta{}^2 - \frac{d\psi_\beta}{dt} - 1 = 0, \quad \psi_\beta(1) = 0$

Let $\psi_\beta = K(t - 1)$

then $\phi_\beta = K^2(t - 1)^2 + K + 1$

and $L(\phi_\beta, \psi_\beta) = \frac{-1}{1260}(K^4 + 420K^3 + 2520K^2 + 1260K + 630)$

The best K is approximately - 0.268, giving $L(\phi_\beta, \psi_\beta) = -0.369$; then

$= (0.268)^2 (t - 1)^2 + 0.732$, which is greater than $-\frac{1}{2}c = -\frac{1}{2}$ for all t, so again the best ϕ_β belongs to U_1, and there is no advantage in using theorem (VI.2.1).

Example 3

This last example illustrates that theorem (VI.2.1) can be used as an alternative to using a functional which includes boundary terms when finding dual extremum principles for a boundary value problem.

Consider the non-linear diffusion problem described by the equations

$$-\frac{d}{d\phi}\left(D(c)\frac{dc}{d\phi}\right) = \frac{1}{2}\phi\frac{dc}{d\phi}, \quad 0 < \phi < \infty \qquad \text{(VI.3.29)}$$

$$c(0) = 1, \quad c(\infty) = 0, \quad \text{where } D(c) \in C^1(0,1).$$

It is shown in (6) that this problem can be written as

$$\frac{1}{2}\phi + \frac{d\psi}{dc} = 0, \quad \frac{D(c)}{\psi} - \frac{d\phi}{dc} = 0, \quad 0 < c < 1 \qquad \text{(VI.3.30)}$$

$$\phi(1) = \psi(0) = 0$$

The two usual methods of obtaining dual extremum principles are as follows:

1) Let $L(\phi,\psi) = \int_0^1 \left\{\frac{1}{4}\phi^2 + D(c)\ \ln|\psi| + \phi\frac{d\psi}{dc}\right\} dc$ (VI.3.31)

where $\phi(1) = \psi(0) = 0 \quad \forall\ (\phi,\psi) \in U_1 \times V_1$,

where $U_1 = \{\phi : \phi \in C^1(0,1),\ \phi(1) = 0\}$ (VI.3.32)

and $V_1 = \{\psi : \psi \in C^1(0,1),\ \psi(0) = 0\}$ (VI.3.33)

It is easily shown that $L(\phi,\psi)$ is a convex/concave saddle functional if $D(c) \geq 0\ \forall\ c$. The gradients are

$$\nabla_\phi L(\phi,\psi) = \frac{1}{2}\phi + \frac{d\psi}{dc}, \qquad \nabla_\psi L(\phi,\psi) = \frac{D(c)}{\psi} - \frac{d\phi}{dc}$$

We then have the dual extremum principles

$L(\phi_\beta,\psi_\beta) \leq L(\phi_e,\psi_e) \leq L(\phi_\alpha,\psi_\alpha)$, where

$\nabla_\phi L(\phi_\alpha,\psi_\beta) = 0, \quad \nabla_\psi L(\phi_\alpha,\psi_\beta) = 0$ and

$(\phi_e, \phi_\alpha, \phi_\beta) \in U_1, \quad (\psi_e, \psi_\alpha, \psi_\beta) \in V_1$

Using this functional, we require $\phi_\beta(1) = \phi_\alpha(1) = 0$,

$\psi_\beta(0) = \psi_\alpha(0) = 0$, which is somewhat restrictive.

2) To decrease the restrictions in the trial functions, we can incorporate boundary conditions in the functional:

Let $L(\phi, \psi) = \int_0^1 \{ \frac{1}{4}\phi^2 + D(c)\,\ln|\psi| \quad + \phi \frac{d}{dc}\psi \}\, dc$

$$+ \phi(0)\,\psi(0) \qquad \text{(VI.3.34)}$$

We have the gradients

$$\nabla_\phi L(\phi, \psi) = \begin{bmatrix} \frac{1}{2}\phi + \frac{d\psi}{dc} \\ 0 \\ \psi(0) \end{bmatrix} \qquad \text{(VI.3.35)}$$

and

$$\nabla_\psi L(\phi, \psi) = \begin{bmatrix} \frac{D(c)}{\psi} - \frac{d\phi}{dc} \\ \phi(1) \\ 0 \end{bmatrix} \qquad \text{(VI.3.36)}$$

This automatically results in the trial functions satisfying:

$$\tfrac{1}{2}\phi_\beta + \frac{d\psi_\beta}{dc} = 0, \qquad \psi_\beta(0) = 0 \qquad \text{(VI.3.37)}$$

and $\frac{D(c)}{\psi_\alpha} - \frac{d\phi_\alpha}{dc} = 0, \qquad \phi_\alpha(1) = 0$ (VI.3.38)

Theorem (VI.2.1) can be used as a third method of obtaining bounds whose trial functions satisfy equations (VI.3.37) and (VI.3.38), without the additional requirements that $\phi_\beta(1) = \psi_\alpha(0) = 0$, but which does not involve boundary terms being incorporated in the functional.

3) Let $L(\phi, \psi) = \int_0^1 \quad \frac{1}{4}\phi^2 + D(c)\,\ln|\psi| \quad + \phi \frac{d\psi}{dc} \}\, dc$ (VI.3.39)

and define the sets U_1, U_2, V_1, V_2 as follows:

$$U_1 = \{ \phi : \phi \in C^1(0,1), \quad \phi(1) = 0 \} \qquad \text{(VI.3.40)}$$

$$U_2 = \{ \phi : \phi \in C^1(0,1) \} \qquad \text{(VI.3.41)}$$

$$V_1 = \{ \psi : \psi \in C^1(0,1), \quad \psi(0) = 0 \} \qquad \text{(VI.3.42)}$$

$$V_2 = \{ \psi : \psi \in C^1(0,1) \} \qquad \text{(VI.3.43)}$$

It can be shown, using equations (VI.2.9) and (VI.2.10) that $L(\phi, \psi)$ is convex in the set U_2 and concave in the set V_2 if $D(c) \geqslant 0$; by theorem (VI.2.1) we have

$L(\phi_\beta, \psi_\beta) \leq L(\phi_e, \psi_e) \leq L(\phi_\alpha, \psi_\alpha)$, where

$\frac{1}{2}\phi_\beta + \dfrac{d\psi_\beta}{dc} = 0$, and $\dfrac{D(c)}{\psi_\alpha} - \dfrac{d\phi_\alpha}{dc} = 0$,

$\phi_\beta \in U_2$, $\psi_\beta \in V_1$, $\phi_\alpha \in U_1$ and $\psi_\alpha \in V_2$; that is $\psi_\beta(0) = 0$ and $\phi_\alpha(1) = 0$ only.

These are, of course, the less restricted conditions on the trial functions obtained using the functional incorporating boundary terms, thus showing that theorem (VI.2.1) provides an alternative method to that incorporating boundary terms.

Ther is obviously no advantage in using theorem (VI.2.1) for this particular problem, but it could be advantageous for problems for which it is difficult to include the boundary terms in the functional.

APPENDIX I

Theorem (III.7.2) : Proof

From the theorem, the relevant equations are

$$A\psi_{2n} + B\phi_{2n+1} + f = 0 \tag{A.I.1}$$

$$A^{x}\phi_{2n} - C\psi_{2n+1} + g = 0 \tag{A.I.2}$$

$$A\psi_{2n+1} + B\phi_{2n+1} + f = 0 \tag{A.I.3}$$

Then, using these equations, the symmetrical form for Q from equation (III.6.10) is

$$Q = -\langle\phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle \tag{A.I.4}$$

$$\lambda_{2n+2} = -\frac{Q}{P} \quad \text{(equation (III.6.14))}$$

$$= \frac{\langle\phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle}{\langle\phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle + \langle\psi_{2n} - \psi_{2n+1},\ C(\psi_{2n} - \psi_{2n+1})\rangle} \tag{A.I.5}$$

(using equation (III.6.9) for P).

Substituting equations (III.6.9), (III.6.11) and (A.I.4) into equation (III.6.15) gives

$$L(\phi_{2n+2}, \psi_{2n+2})_{R} = \frac{\{\langle\phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle\}^2}{2\{\langle\phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle + \langle\psi_{2n} - \psi_{2n+1},\ C(\psi_{2n} - \psi_{2n+1})\rangle\}}$$
$$- \tfrac{1}{2}\langle\phi_{2n+1},\ B\phi_{2n+1}\rangle - \tfrac{1}{2}\langle\psi_{2n+1},\ C\psi_{2n+1}\rangle + \langle\psi_{2n+1},\ g\rangle \tag{A.I.6}$$

$$= \frac{\{\langle\phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle\}^2}{2\{\langle\phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle + \langle\psi_{2n} - \psi_{2n+1},\ C(\psi_{2n} - \psi_{2n+1})\rangle\}} + L(\phi_{2n+1}, \psi_{2n+1}) \tag{A.I.7}$$

(using equation (III.6.5)).

From equation (A.I.7), it is obvious that

$$L(\phi_{2n+2}, \psi_{2n+2})_{R} \geqslant L(\phi_{2n+1}, \psi_{2n+1})_{R}.$$

Also, using equations (A.I.4) and (A.I.6),

$$L(\phi_{2n+2}, \psi_{2n+2})_\beta - L(\phi_{2n}, \psi_{2n})_\beta$$

$$= \frac{\{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}^2}{2\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}}$$

$\geqslant 0$ as B and C are positive-definite operators.

Hence $L(\phi_{2n+2}, \psi_{2n+2})_\beta \geqslant L(\phi_{2n}, \psi_{2n})_\beta$.

APPENDIX II

Derivation of Theorems (III.8.1) and (III.8.2)

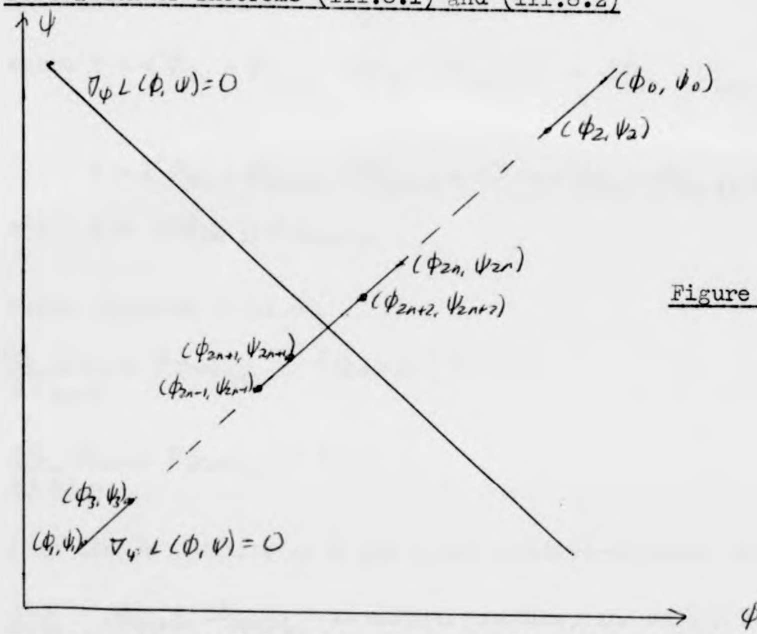


Figure A.II.1

The sequence resulting from the upper bound iterations is

$$L(\phi_e, \psi_e) \leq \quad L(\phi_0, \psi_0)_\alpha, \quad L(\phi_1, \psi_1)_\alpha, \quad L(\phi_2, \psi_2)_\alpha, \quad \ldots\ldots \; L(\phi_{2n}, \psi_{2n})_\alpha,$$
$$L(\phi_{2n+1}, \psi_{2n+1})_\alpha, \quad L(\phi_{2n+2}, \psi_{2n+2})_\alpha \tag{A.II.1}$$

where $\phi_{2n+2} = \lambda_{2n+2}\,(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}$ (A.II.2)

and $\psi_{2n+2} = \lambda_{2n+2}\,(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1} \quad \lambda_{2n+2} \in \mathbb{R}$ (A.II.3)

Hence
$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha = L(\lambda_{2n+2}\,(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1},\ \lambda_{2n+2}\,(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1})_\alpha \tag{A.II.4}$$

For the quadratic functional given by equation (III.2.1), $L(\phi_{2n+2}, \psi_{2n+2})_\alpha$ is given by equation (III.2.8):

$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha = \tfrac{1}{2}\langle \phi_{2n+2}, B\phi_{2n+2}\rangle + \tfrac{1}{2}\langle \psi_{2n+2}, C\psi_{2n+2}\rangle + \langle \phi_{2n+2}, f\rangle$$

Using equations (A.II.2) and (A.II.3), and noting that B and C are symmetric operators,

$$\begin{aligned} L(\phi_{2n+2}, \psi_{2n+2})_\alpha = {} & \frac{\lambda^2_{2n+2}}{2}\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle \\ & + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\} \\ & + \lambda_{2n+2}\{\langle \phi_{2n} - \phi_{2n+1}, B\phi_{2n+1} + f\rangle + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1}\rangle\} \\ & + L(\phi_{2n+1}, \psi_{2n+1})_\alpha \end{aligned} \tag{A.II.5}$$

APPENDIX II

Let $L(\phi_{2n+2}, \psi_{2n+2})_{\alpha} = \frac{\lambda^2_{2n+2}}{2} X + \lambda_{2n+2} Y + Z$ (A.II.6)

where $X = \langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle$ (A.II.7)

$Y = \langle \phi_{2n} - \phi_{2n+1}, B\phi_{2n+1} + f \rangle + \langle \psi_{2n} - \psi_{2n+1}, C\psi_{2n+1} \rangle$ (A.II.8)

and $Z = L(\phi_{2n+1}, \psi_{2n+1})_{\alpha}$ (A.II.9)

Using equation (A.II.6),

$\frac{dL}{d\lambda_{2n+2}} (\phi_{2n+2}, \psi_{2n+2})_{\alpha} = \lambda_{2n+2} X + Y$ (A.II.10)

$\frac{d^2L}{d\lambda^2_{2n+2}} (\phi_{2n+2}, \psi_{2n+2})_{\alpha} = X$ (A.II.11)

X is always positive as B and C are positive-definite operators; hence

$\frac{d^2 L}{d\lambda^2_{2n+2}} (\phi_{2n+2}, \psi_{2n+2})_{\alpha}$ is always positive, and so, by equation (II.8.4),

$L(\phi_{2n+2}, \psi_{2n+2})_{\alpha}$ is convex in λ_{2n+2}. By Lemma (II.10.1), the minimum of $L(\phi_{2n+2}, \psi_{2n+2})_{\alpha}$ occurs when $\frac{dL(\phi_{2n+2}, \psi_{2n+2})_{\alpha}}{d\lambda_{2n+2}} = 0$

that is, when $\lambda_{2n+2} = -\frac{Y}{X}$ (A.II.12)

(As in section III.6, $X = 0$ implies that $Y = 0$, so it is unnecessary to insist that $X \neq 0$).

Substituting equation (A.II.12) into equation (A.II.6) gives

$\text{Min. } L(\phi_{2n+2}, \psi_{2n+2})_{\alpha} = -\frac{Y^2}{2X} + Z$ (A.II.13)

If, in equations (A.II.2) and (A.II.3), λ_{2n+2} is chosen as $-\frac{Y}{X}$, then

$L(\phi_{2n+2}, \psi_{2n+2})_{\alpha} \leq L(\lambda_{2n+2} (\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1}, \lambda_{2n+2} (\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1})_{\alpha}$
$\forall \lambda_{2n+2} \in \mathbb{R}$ (A.II.14)

If λ_{2n+2} is taken as 0 on the right hand side of equation (A.II.16),

$L(\phi_{2n+2}, \psi_{2n+2})_{\alpha} \leq L(\phi_{2n+1}, \psi_{2n+1})_{\alpha}$ (A.II.15)

and if λ_{2n+2} is taken as 1 on the right hand side of equation (A.II.14),

$L(\phi_{2n+2}, \psi_{2n+2})_{\alpha} \leq L(\phi_{2n}, \psi_{2n})_{\alpha}$ (A.II.18)

Hence we have a non-increasing sequence:

$$L(\phi_e, \psi_e) \leq L(\phi_{2n+2}, \psi_{2n+2})_\alpha \leq L(\phi_{2n}, \psi_{2n})_\alpha \leq \;\ldots\ldots$$
$$\ldots\ldots \leq L(\phi_2, \psi_2)_\alpha \leq L(\phi_0, \psi_0)_\alpha \qquad \text{(A.II.17)}$$

The specific iterations in theorems (III.8.1) and (III.8.2) are now proven.

APPENDIX II

Theorem III.8.1 : Proof

From the theorem, the relevant equations are:

$$A^x\phi_{2n} - C\psi_{2n} + g = 0 \qquad (A.II.18)$$

$$A\psi_{2n+1} + B\phi_{2n} + f = 0 \qquad (A.II.19)$$

$$A^x\phi_{2n+1} - C\psi_{2n+1} + g = 0 \qquad (A.II.20)$$

The symmetrical form for Y (equation (A.II.8) is, using the above three equations,

$$Y = -\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle \qquad (A.II.21)$$

Hence, using equations (A.II.12), A.II.7), (A.II.21),

$$\lambda_{2n+2} = \frac{\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle}{\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle} \qquad (A.II.22)$$

and using equations (A.II.13), (A.II.7), (A.II.9) and (A.II.21),

$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha = \frac{-\{\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle\}^2}{2\{\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}} + L(\phi_{2n+1}, \psi_{2n+1})_\alpha \qquad (A.II.23)$$

As B and C are positive-definite operators, it is obvious from equation (A.II.23) that

$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha \leq L(\phi_{2n+1}, \psi_{2n+1})_\alpha$$

Also, using equations (III.2.8), (A.II.8), (A.II.21) and (A.II.23),

$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha - L(\phi_{2n}, \psi_{2n})_\alpha$$

$$= \frac{-\{\langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}^2}{2\{\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}}$$

$$\leq 0$$

Hence $L(\phi_{2n+2}, \psi_{2n+2})_\alpha \leq L(\phi_{2n}, \psi_{2n})_\alpha$.

APPENDIX II

Theorem III.8.2 : Proof

The equations defining the iteration are:

$$A^x\phi_{2n} - C\psi_{2n} + g = 0 \qquad \text{(A.II.24)}$$

$$A\psi_{2n} + B\phi_{2n+1} + f = 0 \qquad \text{(A.II.25)}$$

$$A^x\phi_{2n+1} - C\psi_{2n+1} + g = 0 \qquad \text{(A.II.26)}$$

The symmetrical form for Y, (equation (A.II.8)) is, from the above three equations,

$$Y = -\langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle \qquad \text{(A.II.27)}$$

Therefore, using equations (A.II.12), (A.II.7) and (A.II.27),

$$\lambda_{2n+2} = \frac{\langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}{\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle} \qquad \text{(A.II.28)}$$

and, using equations (A.II.13), (A.II.7), (A.II.9) and (A.II.27),

$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha = \frac{-\{\langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}^2}{2\{\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}} + L(\phi_{2n+1}, \psi_{2n+1})_\alpha \qquad \text{(A.II.29)}$$

$L(\phi_{2n+2}, \psi_{2n+2})_\alpha - L(\phi_{2n+1}, \psi_{2n+1})_\alpha \leq 0$ from equation (A.II.29), as B and C are positive-definite operators.

Using equations (III.2.8), (A.II.8), (A.II.27) and (A.II.29),

$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha - L(\phi_{2n}, \psi_{2n})_\alpha$$

$$= \frac{-\{\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle\}^2}{2\{\langle\phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle\psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle\}}$$

$$\leq 0$$

Hence $L(\phi_{2n+2}, \psi_{2n+2})_\alpha \leq L(\phi_{2n}, \psi_{2n})_\alpha$.

APPENDIX III

Conditions for convergence using the steepest descent method

(i) **Bounded operators - Lower bound iteration**

(a) The equations in theorem (III.7.2) can be written

$$\phi_{2n} = - B^{-1} (A \psi_{2n} + f) \qquad (A.III.1)$$

$$A^x \phi_{2n} - C \psi_{2n+1} + g = 0 \qquad (A.III.2)$$

$$\phi_{2n+1} = - B^{-1} (A \psi_{2n+1} + f) \qquad (A.III.3)$$

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle (\psi_{2n} - \psi_{2n+1})}{\{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle\}} \qquad (A.III.4)$$

Using (A.III.1) and (A.III.3), (A.III.4) becomes

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle (\psi_{2n} - \psi_{2n+1})}{\langle \psi_{2n} - \psi_{2n+1}, (A^x B^{-1} A + C) (\psi_{2n} - \psi_{2n+1}) \rangle} \qquad (A.III.5)$$

Comparing equations (III.9.1) and (A.III.5), we require $C = I$ and

$P \psi_{2n} - G = \psi_{2n} - \psi_{2n+1}$

or $(A^x B^{-1} A + C) \psi_{2n} - g + A^x B^{-1} f = \psi_{2n} - \psi_{2n+1}$

$(A^x B^{-1} A + C) \psi_{2n} - g + A^x B^{-1} f$

$= A^x B^{-1} (A \psi_{2n} + f) + C \psi_{2n} - g$

$= - A^x \phi_{2n} + C \psi_{2n} - g$ from (A.III.1)

$= C(\psi_{2n} - \psi_{2n+1})$ from (A.III.2)

$= \psi_{2n} - \psi_{2n+1}$ as we require $C = I$.

Letting $C = I$ gives $P = (A^x B^{-1} A + I)$.

Therefore the iterations specified in theorem (III.7.2) will converge to the unique solution ψ_e of equation (III.9.6), assuming it exists, provided that the operators A, B and C satisfy the following conditions

(a) $C = I$

(b) $A^x B^{-1} A$ is a linear, self-adjoint operator such that $(A^x B^{-1} A + I)$ is positive-definite.

APPENDIX III

(ii) Bounded operators - Upper bound iteration

The steepest descent iteration, theorem (II.17.1), can be written as: If P is a linear, self-adjoint positive-definite operator, then the iteration

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle P\phi_{2n} - F,\ P\phi_{2n} - F\rangle\ (P\phi_{2n} - F)}{\langle P\phi_{2n} - F,\ P(P\phi_{2n} - F)\rangle} \qquad \text{(A.III.6)}$$

converges to the unique solution ϕ_e of $P\phi_e = F$ (A.III.7)

C^{-1} exists if C is bounded below; and hence equation (III.9.4) can be written

$$\psi_e = C^{-1}\ (A^x\phi_e + g) \qquad \text{(A.III.8)}$$

Substituting equation (A.III.8) into (III.9.3) gives

$$(AC^{-1}A^x + B)\phi_e = -\ (AC^{-1}g + f) \qquad \text{(A.III.9)}$$

or $P\ \phi_e = F$ (A.III.10)

where $P = AC^{-1}A^x + B$ and $F = -\ (AC^{-1}g + f)$ (A.III.11)

Provided the operator P is linear, self-adjoint and positive definite, the iteration given in equation (A.III.6), where $F = -\ (AC^{-1}g + f)$, will converge to the unique solution ϕ_e of equation (A.III.10), assuming that it exists.

In this case, theorem (II.17.1) proves that

$$\lim_{n\to\infty} \|\phi_{2n+2} - \phi_e\| = 0.$$

Now $(\phi_{2n+2}, \psi_{2n+2})$ satisfies the equation

$$A^x\phi_{2n+2} - C\psi_{2n+2} + g = 0 \qquad \text{(A.III.12)}$$

Using equations (III.9.4) and (A.III.12),

$$\psi_{2n+2} - \psi_e = C^{-1}A^x(\phi_{2n+2} - \phi_e)$$

and therefore if $\lim_{n\to\infty} \|\phi_{2n+2} - \phi_e\| = 0$ then

$$\lim_{n\to\infty} \|\phi_{2n+2} - \phi_e\| + \|\psi_{2n+2} - \psi_e\| = 0.$$

The equations in theorem (III.8.1) can be written

$$\psi_{2n} = C^{-1}\ (A^x\phi_{2n} + g) \qquad \text{(A.III.13)}$$

APPENDIX III

$$A\psi_{2n+1} + B\phi_{2n} + f = 0 \tag{A.III.14}$$

$$\psi_{2n+1} = C^{-1}\,(A^x\phi_{2n+1} + g) \tag{A.III.15}$$

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle \psi_{2n} - \psi_{2n+1},\ C(\psi_{2n} - \psi_{2n+1})\rangle (\psi_{2n} - \psi_{2n+1})}{\langle \phi_{2n} - \phi_{2n+1},\ B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1},\ C(\psi_{2n} - \psi_{2n+1})\rangle} \tag{A.III.16}$$

Using equations (A.III.13) and (A.III.15), (A.III.16) becomes

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle \phi_{2n} - \phi_{2n+1},\ AC^{-1}A^x(\phi_{2n} - \phi_{2n+1})\rangle\ (\phi_{2n} - \phi_{2n+1})}{\langle \phi_{2n} - \phi_{2n+1},\ (B + AC^{-1}A^x)(\phi_{2n} - \phi_{2n+1})\rangle} \tag{A.III.17}$$

Comparing equations (A.III.6) and (A.III.17), we require

$AC^{-1}A^x = I$ and $P\phi_{2n} - F = \phi_{2n} - \phi_{2n+1}$

or $(B + AC^{-1}A^x)\phi_{2n} + AC^{-1}g + f = \phi_{2n} - \phi_{2n+1}$

$(B + AC^{-1}A^x)\phi_{2n} + AC^{-1}g + f$

$= AC^{-1}(A^x\phi_{2n} + g) + B\phi_{2n} + f$

$= AC^{-1}A^x(\phi_{2n} - \phi_{2n+1})$ from equations (A.III.13) - (A.III.15)

$= \phi_{2n} - \phi_{2n+1}$ as we require $AC^{-1}A^x = I$

Letting $AC^{-1}A^x = I$ gives $P = B + I$, which is linear, positive-definite and symmetric. P will be self-adjoint if B is a bounded operator.

Hence the iterations specified in theorem (III.8.1) will converge to the unique solution ϕ_e of equation (A.III.10), assuming it exists, provided that the operators A, B and C satisfy the following conditions:

(a) $AC^{-1}A^x = I$

(b) B is a bounded operator.

The iterations in theorem (III.8.2) can be written:

$$\psi_{2n} = C^{-1}(A^x\phi_{2n} + g) \tag{A.III.18}$$

$$A\psi_{2n} + B\phi_{2n+1} + f = 0 \tag{A.III.19}$$

APPENDIX III

$$\psi_{2n+1} = C^{-1}(A^x \phi_{2n+1} + g) \qquad \text{(A.III.20)}$$

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle (\phi_{2n} - \phi_{2n+1})}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle} \qquad \text{(A.III.21)}$$

Using equations (A.III.18) and (A.III.20), (A.III.21) becomes

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle (\phi_{2n} - \phi_{2n+1})}{\langle \phi_{2n} - \phi_{2n+1}, (B + AC^{-1} A^x)(\phi_{2n} - \phi_{2n+1}) \rangle} \qquad \text{(A.III.22)}$$

Comparing equations (A.III.6) and (A.III.22), we require $B = I$ and

$P\phi_{2n} - F = \phi_{2n} - \phi_{2n+1}$

or $(B + AC^{-1} A^x)\phi_{2n} + AC^{-1} g + f = \phi_{2n} - \phi_{2n+1}$

$(B + AC^{-1} A^x)\phi_{2n} + AC^{-1} g + f$

$= AC^{-1} (A^x \phi_{2n} + g) + B\phi_{2n} + f$

$= B(\phi_{2n} - \phi_{2n+1})$ from equations (A.III.18) and (A.III.19)

$= \phi_{2n} - \phi_{2n+1}$, as we require $B = I$.

Letting $B = I$ gives $P = (AC^{-1} A^x + I)$, which is linear, self-adjoint and positive-definite if $AC^{-1} A^x$ is linear and self-adjoint and $(AC^{-1} A^x + I)$ is positive-definite.

Therefore the iterations specified in theorem (III.8.2) will converge to the unique solution ϕ_e of equation (A.III.10), assuming it exists, provided that the operators A, B and C satisfy the following conditions:

(a) $B = I$

(b) $AC^{-1} A^x$ is a linear, self-adjoint operator such that $(AC^{-1} A^x + I)$ is positive-definite.

APPENDIX III

(iii) Unbounded operators - Lower bound iteration

From equation (A.III.5), the iteration in theorem (III.7.2) is

$$\psi_{2n+2} = \psi_{2n} - \frac{\langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle \; (\psi_{2n} - \psi_{2n+1})}{\langle \psi_{2n} - \psi_{2n+1}, (A^x B^{-1} A + C) \; (\psi_{2n} - \psi_{2n+1}) \rangle} \tag{A.III.23}$$

Comparing equations (III.9.2) and (A.III.23), we define Z_{2n} and Q as

$$Z_{2n} = \psi_{2n} - \psi_{2n+1} \tag{A.III.24}$$

$$Q = C \tag{A.III.25}$$

To satisfy equation (III.9.22) we require that the iteration equations (A.III.1) - (A.III.3) are equivalent to

$$C(\psi_{2n} - \psi_{2n+1}) = (A^x B^{-1} A + C)\psi_{2n} - g + A^x B^{-1} f$$

$$\text{or } A^x B^{-1} (A\,\psi_{2n} + f) + C\,\psi_{2n+1} - g = 0 \tag{A.III.26}$$

From equations (A.III.1) and (A.III.2),

$A^x B^{-1} (A\,\psi_{2n} + f) = - A^x \phi_{2n} = - (C\,\psi_{2n+1} - g)$; hence the iteration equations (A.III.1) - (A.III.3) are equivalent to equation (III.9.22). Therefore, the iteration given in theorem (III.7.2) will converge to the solution ψ_e of equation (III.9.29), assuming it exists, provided

(a) $A^x B^{-1} A$ and C are closed, symmetric operators with dense domains in a Hilbert space H, with $D(C) \subseteq D(A^x B^{-1} A + C)$;

(b) There exists a real number γ such that

$$\langle x, Cx \rangle \geq \gamma^2 \langle x,x \rangle \quad \forall\, x \in D(C) \tag{A.III.27}$$

(c) There exists real numbers m and M such that

$$(m - 1) \langle x, Cx \rangle \leq \langle x, A^x B^{-1} Ax \rangle \leq (M - 1) \langle x, Cx \rangle \quad \forall\, x \in D(C) \tag{A.III.28}$$

(iv) Unbounded operators - Upper bound iterations

By theorem (II.18.1), if P and Q are two closed, symmetric operators with dense domains in a Hilbert space H, $D(Q) \subseteq D(P)$, and there exist real numbers m, M and γ such that

(a) $\langle x, Qx \rangle \geq \gamma^2 \langle x, x \rangle \; \forall x \in D(Q)$ and (A.III.29)

(b) $m \langle x, Qx \rangle \leq \langle x, Px \rangle \leq M \langle x, Qx \rangle \; \forall x \in D(Q)$ (A.III.30)

then the iteration

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle Z_{2n}, Q Z_{2n} \rangle}{\langle Z_{2n}, P Z_{2n} \rangle} Z_{2n} \tag{A.III.31}$$

with Z_{2n} given by $Q Z_{2n} = P\phi_{2n} - F$ (A.III.32)

converges to the unique solution ϕ_e of $P\phi_e = F$. (A.III.33)

From equation (A.III.11) we have

$P = AC^{-1} A^x + B$ and $F = -(AC^{-1} g + f)$ (A.III.34)

From equation (A.III.18), the iteration in theorem (III.8.1) is

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle \phi_{2n} - \phi_{2n+1}, AC^{-1} A^x (\phi_{2n} - \phi_{2n+1}) \rangle (\phi_{2n} - \phi_{2n+1})}{\langle \phi_{2n} - \phi_{2n+1}, (B + AC^{-1} A^x) (\phi_{2n} - \phi_{2n+1}) \rangle} \tag{A.III.35}$$

Comparing equations (A.III.31) and (A.III.35) we define Z_{2n} and Q as

$Z_{2n} = \phi_{2n} - \phi_{2n+1}$ (A.III.36)

$Q = AC^{-1} A^x$ (A.III.37)

To satisfy equation (A.III.32) we require that the iteration equations (A.III.13) - (A.III.15) are equivalent to

$AC^{-1} A^x (\phi_{2n} - \phi_{2n+1}) = (B + AC^{-1} A^x)\phi_{2n} + AC^{-1} g + f$

or $AC^{-1} (A^x \phi_{2n+1} + g) + B\phi_{2n} + f = 0$ (A.III.38)

From equations (A.III.14) and (A.III.15),

$AC^{-1} (A^x\phi_{2n+1} + g) = A\psi_{2n+1} = -(B\phi_{2n} + f)$ and hence the iteration equations (A.III.13) - (A.III.15) are equivalent to equation (A.III.32). Therefore, the iteration given in theorem (III.8.1) will converge to the solution ϕ_e of

$(AC^{-1} A^x + B) \phi_e = -(AC^{-1} g + f)$, (A.III.39)

assuming it exists, provided

APPENDIX III

(a) $AC^{-1}A^x$ and B are closed, symmetric operators with dense domains in a Hilbert space H, with $D(AC^{-1}A^x) \subseteq D(AC^{-1}A^x + B)$.

(b) There exists a real number γ such that

$$\langle x, AC^{-1}A^x x\rangle \geq \gamma^2 \langle x, x\rangle \quad \forall x \in D(AC^{-1}A^x) \tag{A.III.40}$$

(c) There exist real numbers m and M such that

$$(m-1)\langle x, AC^{-1}A^x x\rangle \leq \langle x, Bx\rangle \leq (M-1)\langle x, AC^{-1}A^x x\rangle \quad \forall x \in D(AC^{-1}A^x) \tag{A.III.41}$$

From equation (A.III.21) the iteration in theorem (III.8.2) is

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, (B + AC^{-1}A^x)(\phi_{2n} - \phi_{2n+1})\rangle}(\phi_{2n} - \phi_{2n+1}) \tag{A.III.42}$$

Comparing equations (A.III.31) and (A.III.42) we define Z_{2n} and Q as

$$Z_{2n} = \phi_{2n} - \phi_{2n+1} \tag{A.III.43}$$

$$Q = B \tag{A.III.44}$$

To satisfy equation (A.III.32) we require that the iteration equations (A.III.18) - (A.III.20) are equivalent to

$$B(\phi_{2n} - \phi_{2n+1}) = (B + AC^{-1}A^x)\phi_{2n} + AC^{-1}g + f$$

$$\text{or } AC^{-1}(A^x\phi_{2n} + g) + B\phi_{2n+1} + f = 0 \tag{A.III.45}$$

From equations (A.III.18) and (A.III.19),

$AC^{-1}(A^x\phi_{2n} + g) = A\phi_{2n} = -(B\phi_{2n+1} + f)$ and hence the iteration equations (A.III.18) - (A.III.20) are equivalent to equation (A.III.32).

Therefore, the iteration given in theorem (III.8.2) will converge to the solution ϕ_e of equation (A.III.39), assuming it exists, provided

(a) $AC^{-1}A^x$ and B are closed, symmetric operators with dense domains on a Hilbert space H, with $D(B) \subseteq D(B + AC^{-1}A^x)$.

(b) There exists a real number γ such that

$$\langle x, Bx\rangle \geq \gamma^2 \langle x, x\rangle \quad \forall x \in D(B) \tag{A.III.46}$$

(c) There exist real numbers m and M such that

$$(m-1)\langle x, Bx\rangle \leq \langle x, AC^{-1}A^x x\rangle \leq (M-1)\langle x, Bx\rangle \quad \forall x \in D(B). \tag{A.III.47}$$

APPENDIX IV

Iterations specified by equations (III.9.43)

From equations (III.9.40) and (III.9.41),

$$K\phi(x) = \int_0^x y\,\phi(y)\,dy + \int_x^1 x\,\phi(y)\,dy$$

Also $$K(x^n) = \frac{x}{n+1} - \frac{x^{n+2}}{(n+1)(n+2)} \qquad \text{(A.IV.1)}$$

Let $\phi_0 = 0$

$L(\phi_0, \psi_0)_\alpha = 0$

$\phi_1 = x^2 - 2x - K\phi_0 = x^2 - 2x$

$$L(\phi_0, \psi_1) = -\tfrac{1}{2}\int_0^1 (x^2-2x)^2\,dx = -\frac{4}{15} = -0.26$$

$$\lambda_2 = \frac{\int_0^1 (x^2-2x)\,K\,(x^2-2x)\,dx}{\int_0^1 (x^2-2x)\,(K+I)\,(x^2-2x)\,dx} = \frac{17}{59}$$

$$\phi_2 = \frac{42}{59}(x^2-2x)$$

$L(\phi_2, \psi_2)_\alpha = -0.1898305$

$$\phi_3 = x^2 - 2x - K\left(\frac{42}{59}(x^2-2x)\right) = \frac{1}{118}(-180x + 118x^2 - 28x^3 + 7x^4)$$

$$L(\phi_2, \psi_3)_\beta = -\tfrac{1}{2}\int_0^1 (\phi_2\,K\phi_2 + \phi_3^2)\,dx = -0.18985604$$

$$\lambda_4 = \frac{\int_0^1 (\phi_2-\phi_3)\,K\,(\phi_2-\phi_3)\,dx}{\int_0^1 (\phi_2-\phi_3)\,(K+I)\,(\phi_2-\phi_3)\,dx} = \frac{101}{2411}$$

$$\phi_4 = \lambda_4(\phi_2-\phi_3) + \phi_3 = \frac{1}{142249}(-216384x + 140532x^2 - 32340x^3 + 8085x^4)$$

$$L(\phi_4, \psi_4)_\alpha = \tfrac{1}{2}\int_0^1 \phi_4\,((K+I)\phi_4 + 2(2x-x^2))\,dx$$

$$= -0.18985497$$

Using the method in section II.15, the equivalent differential equation is

$\phi''(x) - \phi(x) = 2$, $\phi(0) = \phi'(1) = 0$, which has the solution

APPENDIX IV

$\phi_e(x) = 2 \cosh x - 2 \tanh 1 \sinh x - 2.$

Then $L(\phi_e, \psi_e) = \frac{4}{3} - 2 \tanh 1 = -0.18985497.$

The first four terms in $\phi_e(x)$ are:

$$- 2 \tanh 1 \, x + x^2 - \frac{\tanh 1}{3} x^3 + \frac{x^4}{12}$$

$$= -1.523188312x + x^2 - 0.253864718x^3 + 0.0833x^4$$

APPENDIX IV

Iterations specified by equations (III.9.52)

Let $\phi_o = 0$: Then $\psi_o = 0$

$L(\phi_o, \psi_o)_\alpha = 0$

$\frac{d\psi_1}{dt} = -2$ giving $\psi_1 = -2t + c$; $\psi_1(1) = -2 + c = 0$, so $c = 2$

$\psi_1 = 2 - 2t$

$L(\phi_o, \psi_1)_\beta = -\frac{1}{2}\int_o^1 (2 - 2t)^2\,dt = -\frac{2}{3}$

$\frac{d\phi_1}{dt} = -\psi_1 = 2t - 2$; $\phi_1 = t^2 - 2t + c$ $\quad \phi_1(0) = c = 0$

$\phi_1 = t^2 - 2t$

$$\lambda_2 = \frac{\int_o^1 (t^2 - 2t)^2\,dt}{\int_o^1 \left\{(t^2 - 2t)^2 + (2 - 2t)^2\right\} dt} = \frac{2}{7}$$

$\phi_2 = \frac{5}{7}(t^2 - 2t)$

$\psi_2 = \frac{5}{7}(2 - 2t)$

$$L(\phi_2, \psi_2)_\alpha = \frac{1}{2}\int_o^1 \left\{(\frac{5}{7}(t^2 - 2t)^2 + (\frac{5}{7}(2 - 2t)^2 + 4 \times \frac{5}{7}(t^2 - 2t)\right\} dt$$

$= -\ 0.476190476$

$\frac{d\psi_3}{dt} = -2 - \frac{5}{7}(t^2 - 2t)$ giving $\psi_1 = -2t - \frac{5}{7}\ (\frac{t^3}{3} - t^2) + c$

$\psi_1(1) = -2 - \frac{5}{7}(\frac{1}{3} - 1) + c = 0$ so $c = \frac{32}{21}$

$\psi_3 = -2t + \frac{5}{7}t^2 - \frac{5}{21}t^3 + \frac{32}{21}$

$$L(\phi_2, \psi_3)_\beta = -\frac{1}{2}\int_o^1 (\frac{5}{7}(t^2 - 2t))^2 + (-2t + \frac{5}{7}t^2 - \frac{5}{21}t^3 + \frac{32}{21})^2\ )\,dt$$

$= -\ 0.476838354$

$\frac{d\phi_3}{dt} = 2t - \frac{5}{7}t^2 + \frac{5}{21}t^3 - \frac{32}{21}$

$\phi_3 = t^2 - \frac{5}{21}t^3 + \frac{5}{84}t^4 - \frac{32}{21}t$

APPENDIX IV

$$\lambda_4 = \frac{448883}{449063}$$

$$\phi_4 = \frac{1}{9430323}\,(-\,13472250t + 6737025t^2 - 900t^3 + 225t^4)$$

$$\psi_4 = \frac{1}{9430323}\,(13472250 - 13474050t + 2700t^2 - 900t^3)$$

$$L(\phi_4, \psi_4)_\alpha = -0.47619099$$

From the previous page, the solution of the problem is

$\phi_e(t) = 2 \cosh t - 2 \tanh 1 \sinh t - 2$; then $L(\phi_e, \psi_e)$

$= \int_0^1 \phi_e(t)\, dt = -0.47681168$

APPENDIX V

(i) Iterations given by equation (III.11.49)

Let $\phi_o = 0 \qquad L(\phi_o, u_o)_\alpha = 0$

$$\frac{1}{r}\frac{\partial}{\partial r}(r\,u_1) = \phi_o + 1 = 1 \qquad \underline{u}_1 = \frac{r}{2}\,\underline{e}_r$$

$$L(\phi_o, \underline{u}_1)_\beta = -0.19634954$$

$$\frac{\partial\phi_1}{\partial r} = \frac{r}{2} \qquad \phi_1 = \frac{r^2}{4} + c \qquad \phi_1(1) = \frac{1}{4} + c = 0 \qquad c = -\frac{1}{4} \qquad \phi_1 = -\frac{1}{4}(1-r^2)$$

$$L(\phi_1, \underline{u}_1)_\alpha = -0.163624617$$

$$\frac{1}{r}\frac{\partial}{\partial r}(r\,u_2) = -\frac{1}{4}(1 - r^2) + 1 \qquad \underline{u}_2 = \frac{1}{16}(6r + r^3)\,\underline{e}_r$$

$$L(\phi_1, \underline{u}_2)_\beta = -0.169249213$$

$$\frac{\partial\phi_2}{\partial r} = \frac{1}{16}(6r + r^3) \qquad \phi_2 = -\frac{1}{64}(13 - 12r^2 - r^4)$$

$$L(\phi_2, \underline{u}_2)_\alpha = -0.168277692$$

$$\frac{1}{r}\frac{\partial}{\partial r}(r\,u_3) = 1 - \frac{1}{64}(13 - 12r^2 - r^4) \qquad \underline{u}_3 = \frac{1}{384}(153r + 18r^3 + r^5)\,\underline{e}_r$$

$$L(\phi_2, \underline{u}_3)_\beta = -0.168445649$$

$$\frac{\partial\phi_3}{\partial r} = \frac{1}{384}(153r + 18r^3 + r^5) \qquad \phi_3 = -\frac{1}{2304}(487 - 459r^2 - 27r^4 - r^6)$$

$$L(\phi_3, \underline{u}_3)_\alpha = -0.168416608$$

Iterations given by equations (III.11.50)

Let $\phi_o = 0 \qquad L(\phi_o, \underline{u}_o)_\alpha = 0$

$$\frac{1}{r}\frac{\partial}{\partial r}(r\,u_1) = \phi_o + 1 \qquad \underline{u}_1 = \frac{r}{2}\,\underline{e}_r$$

$$L(\phi_o, \underline{u}_1)_\beta = -0.19634954$$

$$\phi_1 = -\frac{1}{4}(1 - r^2)$$

$$\lambda_2 = \frac{\int_0^1 \left(\frac{1}{4}(1 - r^2)\right)^2 r\,dr}{\int_0^1 \left\{\left(\frac{1}{4}(1 - r^2)\right)^2 + \left(\frac{r}{2}\right)^2\right\} r\,dr} = \frac{1}{7}$$

APPENDIX V

$$\phi_2 = -\frac{3}{14}(1 - r^2) \qquad \underline{u}_2 = \frac{3}{7}\, r \;\underline{e}r$$

(ii) $$L(\phi_2, \underline{u}_2)_\alpha = \pi\int_0^1 \left\{(-\frac{3}{14}(1 - r^2))^2 + (\frac{3}{7}\,r)^2 - \frac{3}{7}(1 - r^2)\right\} r\,dr$$

$$= \frac{-3\pi}{56} = -0.168299606$$

$$\frac{1}{r}\frac{\partial}{\partial r}(r\,u_3) = \phi_2 + 1 = \frac{1}{14}(11 + 3r^2) \qquad \underline{u}_3 = \frac{1}{56}(22r + 3r^3)\,\underline{e}r$$

$$L(\phi_2, \underline{u}_3)_\beta = -\pi\int_0^1 \left\{(-\frac{3}{14}(1 - r^2))^2 + (\frac{1}{56}(22r + 3r^3))^2\right\} r\,dr$$

$$= -\frac{1345}{25088} = -0.168424829$$

$$\frac{\partial\phi_3}{\partial r} = \frac{1}{56}(22r + 3r^3) \qquad \phi_3 = -\frac{1}{224}(47 - 44r^2 - 3r^4)$$

$$\lambda_4 = \frac{\int_0^1 (\frac{1}{224}(-1 + 4r^2 - 3r^4))^2\, r\,dr}{\int_0^1 (\frac{1}{224}(-1 + 4r^2 - 3r^4))^2 + (\frac{1}{224}(8r - 12r^3))^2 \quad r\,dr}$$

$$= \frac{1}{31}$$

$$\phi_4 = \frac{-1}{3472}(729 - 684r^2 - 45r^4)$$

$$\underline{u}_4 = \frac{1}{868}(342r + 45r^3)$$

$$L(\phi_4, \underline{u}_4) = \int_0^1 (\frac{1}{3472}(729 - 684r^2 - 45r^4))^2 + (\frac{1}{3472}(1368r + 180r^3))^2 - \frac{2}{3472}(729 - 684r^2 - 45r^4) \quad r\,dr$$

$$= -0.16842078$$

$$\frac{1}{r}\frac{\partial}{\partial r}(r\,u_5) = \quad 4 + 1 = \frac{1}{3472}(2743 + 684r^2 + 45r^4)$$

$$\underline{u}_5 = \frac{1}{6944}(2743r + 342r^3 + 15r^5)$$

APPENDIX V

$$\frac{\partial \phi_5}{\partial r} = \frac{1}{6944}\left(2743r + 342r^3 + 15r^5\right)$$

$$\phi_5 = \frac{-1}{13888}\left(2919 - 2743r^2 - 171r^4 - 5r^6\right)$$

$$L(\phi_4, u_5)_\alpha = -0.168420903$$

$$\lambda_6 = \frac{\int_0^1 \left(\frac{1}{13888}\left(3 - 7r^2 + 9r^4 - 5r^6\right)\right)^2 r\,dr}{\int_0^1 \left(\frac{1}{13888}\left(3 - 7r^2 + 9r^4 - 5r^6\right)\right)^2 + \left(\frac{1}{13888}\left(-14r + 36r^3 - 30r^5\right)\right)^2 \; r\,dr}$$

$$= \frac{221}{1691}$$

$$\phi_6 = \frac{1}{23484608}\left(-4935366 + 4636866r^2 + 291150r^4 + 7350r^6\right)$$

$$u_6 = \frac{1}{23484608}\left(9273732r + 1164600r^3 + 44100r^5\right)$$

$$L(\phi_6, u_6)_\alpha = -0.16842088$$

(iii) Calculation of $L(\phi_e, u_e)$ for the example in Section III.11

From equations (III.11.25), (III.11.27) and (III.11.57),

$$L(\phi_e, u_e) = 2\pi \int_0^1 \left\{ \frac{1}{2}\left(\frac{\partial}{\partial r}\left(-1 + \frac{I_0(r)}{I_0(1)}\right)\right)^2 + \frac{1}{2}\left(-1 + \frac{I_0(r)}{I_0(1)}\right)^2 - 1 + \frac{I_0(r)}{I_0(1)} \right\} r\,dr$$

$$= 2\pi \int_0^1 \left\{ \frac{1}{2}\left(\frac{I_1(r)}{I_0(1)}\right)^2 + \frac{1}{2} - \frac{I_0(r)}{I_0(1)} + \frac{1}{2}\left(\frac{I_0(r)}{I_0(1)}\right)^2 - 1 + \frac{I_0(r)}{I_0(1)} \right\} r\,dr$$

as $\frac{\partial}{\partial r} I_0(r) = I_1(r)$, from page 961 of (40)

$$L(\phi_e, u_e) = \pi \int_0^1 \left\{ r\left(\frac{I_1(r)}{I_0(1)}\right)^2 + r\left(\frac{I_0(r)}{I_0(1)}\right)^2 - r \right\} dr \tag{A.V.1}$$

APPENDIX V

Now,

$$I_0(r) = J_0(e^{\frac{i\pi}{2}} r)$$

$$I_1(r) = e^{-\frac{i\pi}{2}} J_1(e^{\frac{i\pi}{2}} r) \qquad (A.V.2)$$

$$I_2(r) = -J_2(e^{\frac{i\pi}{2}} r)$$

(Page 952 of (40))

$$J_{-1}(r) = -J_1(r) \qquad (A.V.3)$$

(Page 951 of (40))

and

$$\int r\left(J_0(e^{\frac{i\pi}{2}} r)\right)^2 dr = \frac{r^2}{2}\left\{\left(J_0(e^{\frac{i\pi}{2}} r)\right)^2 - J_{-1}(e^{\frac{i\pi}{2}} r)\, J_1(e^{\frac{i\pi}{2}} r)\right\} \qquad (A.V.4)$$

$$\int r\left(J_1(e^{\frac{i\pi}{2}} r)\right)^2 dr = \frac{r^2}{2}\left\{\left(J_1(e^{\frac{i\pi}{2}} r)\right)^2 - J_0(e^{\frac{i\pi}{2}} r)\, J_2(e^{\frac{i\pi}{2}} r)\right\} \qquad (A.V.5)$$

(Page 634 of (40))

Therefore

$$\int_0^1 r\left(I_0(r)\right)^2 dr = \int_0^1 r\left(J_0(e^{\frac{i\pi}{2}} r)\right)^2 dr$$

$$= \left(\frac{r^2}{2}\left\{\left(J_0(e^{\frac{i\pi}{2}} r)\right)^2 + \left(J_1(e^{\frac{i\pi}{2}} r)\right)^2\right)\right)_0^1$$

$$= \left(\frac{r^2}{2}\left\{\left(I_0(r)\right)^2 + e^{\pi i}\left(I_1(r)\right)^2\right\}\right)_0^1$$

$$= \left(\frac{r^2}{2}\left\{\left(I_0(r)\right)^2 - \left(I_1(r)\right)^2\right\}\right)_0^1$$

$$= \tfrac{1}{2}\left(\left(I_0(1)\right)^2 - \left(I_1(1)\right)^2\right)$$

and

$$\int_0^1 r\left(I_1(r)\right)^2 dr = e^{-\pi i}\int_0^1 r\left(J_1(e^{\frac{i\pi}{2}} r)\right)^2 dr$$

$$= -\int_0^1 r\left(J_1(e^{\frac{i\pi}{2}} r)\right)^2 dr$$

$$= -\left(\frac{r^2}{2}\left\{\left(J_1(e^{\frac{i\pi}{2}} r)\right)^2 - J_0(e^{\frac{i\pi}{2}} r)\, J_2(e^{\frac{i\pi}{2}} r)\right\}\right)_0^1$$

$$= -\left(\frac{r^2}{2}\left\{ e^{\pi i}\left(I_1(r)\right)^2 + I_0(r)\, I_2(r)\right\}\right)_0^1$$

$$= \left(\frac{r^2}{2}\left\{\left(I_1(r)\right)^2 - I_0(r)\, I_2(r)\right\}\right)_0^1$$

$$= \tfrac{1}{2}\left(\left(I_1(1)\right)^2 - I_0(1)\, I_2(1)\right)$$

Hence

$$L(\phi_e, u_e) = \pi\left(\tfrac{1}{2}\,\frac{\left(I_1(1)\right)^2 - I_0(1)\, I_2(1) + \left(I_0(1)\right)^2 - \left(I_1(1)\right)^2}{\left(I_0(1)\right)^2} - \tfrac{1}{2}\right)$$

That is

$$L(\phi_e, u_e) = -\frac{\pi}{2}\,\frac{I_2(1)}{I_0(1)}$$

Finally, from page 416 of (40)

$I_0(1) = e^1\,(0.465759608)$

$I_2(1) = 0.135747670$

giving $L(\phi_e, u_e) = -0.168420889$

APPENDIX VI

Derivation of Optimising Iteration D, section (IV.5)

Let $\phi_1 = \phi_4 = \phi_{2n}$, $\phi_2 = \phi_{2n+1}$ and $\psi_1 = \psi_2 = \psi_3 = \psi_{2n}$;

then the defining equations are

$$A^x \phi_{2n} - C\psi_{2n} + g = 0 \qquad (A.VI.1)$$

$$A\psi_{2n} + B_N\phi_{2n} + B_M\phi_{2n+1} + f = 0 \qquad (A.VI.2)$$

$$A^x \phi_{2n+1} - C\psi_{2n+1} + g = 0 \qquad (A.VI.3)$$

$$\phi_{2n+2} = \lambda_{2n+2}(\phi_{2n} - \phi_{2n+1}) + \phi_{2n+1} \qquad (A.VI.4)$$

$$\psi_{2n+2} = \lambda_{2n+2}(\psi_{2n} - \psi_{2n+1}) + \psi_{2n+1} \qquad (A.VI.5)$$

where λ_{2n+2} is chosen to minimise the upper bound

$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha = \tfrac{1}{2}\langle \phi_{2n+2}, B\phi_{2n+2}\rangle + \tfrac{1}{2}\langle \psi_{2n+2}, C\psi_{2n+2}\rangle + \langle \phi_{2n+2}, f\rangle \qquad (A.VI.6)$$

Substituting equations (A.VI.4) and (A.VI.5) into (A.VI.7) gives

$$L(\phi_{2n+2}, \psi_{2n+2})_\alpha = \frac{X}{2}\left(\lambda_{2n+2} + \frac{Y}{X}\right)^2 + Z - \frac{Y^2}{2X} \qquad (A.VI.7)$$

where $X = \langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle$ (A.VI.8)

$$Y = \langle \phi_{2n} - \phi_{2n+1}, B\phi_{2n+1} + f\rangle + \langle \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle \qquad (A.VI.9)$$

and $Z = L(\phi_{2n+1}, \psi_{2n+1})_\alpha$ (A.VI.10)

$L(\phi_{2n+2}, \psi_{2n+2})_\alpha$ takes its minimum when $\lambda_{2n+2} = -\frac{Y}{X}$

Using equations (A.VI.1) to (A.VI.3), Y can be written

$$Y = -\langle \phi_{2n} - \phi_{2n+1}, B_N(\phi_{2n} - \phi_{2n+1})\rangle - \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle$$

and hence the best λ_{2n+2} is

$$\lambda_{2n+2} = \frac{\langle \phi_{2n} - \phi_{2n+1}, B_N(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle}$$

$$= 1 - \frac{\langle \phi_{2n} - \phi_{2n+1}, B_M(\phi_{2n} - \phi_{2n+1})\rangle}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1})\rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1})\rangle} \qquad (A.VI.11)$$

APPENDIX VI

For convergence, using theorem (II.17.1), we need to find a linear, self-adjoint, positive-definite operator P such that the equations in the iterative scheme are equivalent to

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle Z_{2n}, Z_{2n} \rangle Z_{2n}}{\langle Z_{2n}, PZ_{2n} \rangle} \qquad \text{(A.VI.12)}$$

where $Z_{2n} = P\phi_{2n} - F$ and F satisfies $P\phi_e = F$ (A.VI.13)

Then $\lim_{n \to \infty} \{ \| \phi_{2n+2} - \phi_e \| + \| \psi_{2n+2} - \psi_e \| \} = 0$

From equations (A.VI.4) and (A.VI.11), the iteration schemes involving ϕ_{2n+2} can be written

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle \phi_{2n} - \phi_{2n+1}, B_M(\phi_{2n} - \phi_{2n+1}) \rangle (\phi_{2n} - \phi_{2n+1})}{\langle \phi_{2n} - \phi_{2n+1}, B(\phi_{2n} - \phi_{2n+1}) \rangle + \langle \psi_{2n} - \psi_{2n+1}, C(\psi_{2n} - \psi_{2n+1}) \rangle}$$

or, using equations (A.VI.1) and (A.VI.3),

$$\phi_{2n+2} = \phi_{2n} - \frac{\langle \phi_{2n} - \phi_{2n+1}, B_M(\phi_{2n} - \phi_{2n+1}) \rangle (\phi_{2n} - \phi_{2n+1})}{\langle \phi_{2n} - \phi_{2n+1}, (B + AC^{-1} A^x)(\phi_{2n} - \phi_{2n+1}) \rangle} \qquad \text{(A.VI.14)}$$

If we take $B_M = \frac{1}{m^2} I$, where m is a real, non-zero number, then the last equation becomes

$$\phi_{2n+2} = \phi_n - \frac{\langle \phi_{2n} - \phi_{2n+1}, \phi_{2n} - \phi_{2n+1} \rangle (\phi_{2n} - \phi_{2n+1})}{\langle \phi_{2n} - \phi_{2n+1}, m^2(B + AC^{-1} A^x)(\phi_{2n} - \phi_{2n+1}) \rangle} \qquad \text{A.VI.15)}$$

Comparing equations (A.VI.12) and (A.VI.15) we therefore need

$Z_{2n} = \phi_{2n} - \phi_{2n+1}$ and $P = m^2(B + AC^{-1} A^x)$ (A.VI.16)

From equations (IV.5.4) and (IV.5.5),

$(B + AC^{-1} A^x)\phi_e = - (AC^{-1} g + f)$

or $m^2 (B + AC^{-1} A^x)\phi_e = - m^2 (AC^{-1} g + f)$ (A.VI.17)

Then, from equation (A.VI.13), we must take

$F = - m^2 (AC^{-1} g + f)$ (A.VI.18)

and, for convergence, we require that the iteration equations (A.VI.1) to (A.VI.3) are equivalent to

APPENDIX VI

$$\phi_{2n} - \phi_{2n+1} = m^2 (B + AC^{-1} A^x)\phi_{2n} + m^2 (AC^{-1} g + f) \qquad \text{(A.VI.19)}$$

Using equations (A.VI.1) to (A.VI.3),

$$B_M (\phi_{2n} - \phi_{2n+1}) = B_M \phi_{2n} + f + A\psi_{2n} + B_N \phi_{2n}$$
$$= B\phi_{2n} + f + A(C^{-1} (A^x\phi_{2n} + g))$$
$$= (B + AC^{-1} A^x)\phi_{2n} + (AC^{-1} g + f);$$

as we are taking $B_M = \frac{1}{m^2} I$,

$\phi_{2n} - \phi_{2n+1} = m^2 (B + AC^{-1} A^x)\phi_{2n} + m^2 (AC^{-1} g + f)$, which is the same as equation (A.VI.19); hence the iteration given by equations (A.VI.1) to (A.VI.5) and (A.VI.11) will converge provided $B_M = \frac{1}{m^2} I$ and $P = m^2(B + AC^{-1}A^x)$ is linear, self-adjoint and positive definite.

REFERENCES

(1)	Ambramowitz, A. and Stegun, I.A.	Handbook of Mathematical Functions, (1965), Dover Publishing Inc., New York
(2)	Anderson, N. and Arthurs, A.M.	Proc. Camb. Phil. Soc, (1970), 68, 173-177 Complementary Principles for $\phi^{iv} = f(\phi)$
(3)	Anderson, N. and Arthurs, A.M.	J. Engng. Maths. (1972), 6 (4), 331-339 Extremum Principles for some non-linear heat transfer problems.
(4)	Anderson, N. and Arthurs, A.M.	Bull. Math. Biol. (1976), 38, 83-85 Extremum Principles and the elastic and fluid mechanical behaviour of the human eye.
(5)	Anderson, N. and Arthurs, A.M.	J. Math. Anal. Appl. (1976), 56, 736-740 Variational Principles for Complex Equations.
(6)	Anderson, N. and Arthurs, A.M.	Q. Appl. Maths (1977), 188-190 Dual Extremum Principles for a Non-linear Diffusion Problem.
(7)	Arthurs, A.M.	Complementary Variational Principles, (1970) Clarendon Press, Oxford
(8)	Arthurs, A.M.	J. Math. Anal. Appl. (1973), 41, 781-795 Dual Extremum Principles and Error Bounds for a class of boundary value problems.
(9)	Arthurs, A.M.	Complementary Variational Principles, 2nd Edition, (1980), Clarendon Press, Oxford
(10)	Arthurs, A.M. and Anderson, N.	Il Nuovo Cimento, (1968), 55 B, 566-569 A problem in the theory of plasmas.
(11)	Arthurs, A.M. and Coles, C.W.	Proc. Camb. Phil. Soc, (1972), 72, 229-232 Complementary variational principles for a class of differential equations.
(12)	Arthurs, A.M. and Duggan, R.C.	Q. App. Maths (1975), 304-307 Modified Complementary Variational Principles for a Boundary Value Problem in the Twisting of Ring Sectors.

(13) Arthurs, A.M. and Duggan, R.C. J. Math. Anal. Appl. (1977), 59, 365-375 Dual Extremum Principles and the Hypercircle for Biharmonic Problems.

(14) Arthurs, A.M. and Jones, M.G. J. Math. Anal. Appl. (1976), 54, 840-865 On Variational Principles for Linear Initial Value Problems.

(15) Bailey, P. and Smith, P. Math. Methods. App. Sci. (In Press) Application of Convex Analysis and Comparison Functionals to Extremum Principles.

(16) Barnsley, M.F. Phys. Lett. (1975), 53 A(2), 124-126 Pointwise bounds for eigenfunctions of one-electron systems.

(17) Barnsley, M.F. and Robinson, P.D. J. Inst. Math. Appl. (1974), 14, 229-249 Dual variational principles and Padé-type approximations.

(18) Barnsley, M.F. and Robinson, P.D. Proc. Roy. Soc. (1974), A338, 527-583 Bivariational Bounds

(19) Barnsley, M.F. and Robinson, P.D. Proc. Roy. Soc. Edinburgh (1975/76), 75 A(9), 109-118. Bivariational Bounds associated with non-self-adjoint linear operators.

(20) Barnsley, M.F. and Robinson, P.D. J. Math. Anal. Appl. (1976), 56, 172-184 Bivariational Bounds on Solutions of Two Point Boundary Value Problems.

(21) Barnsley, M.F. and Robinson, P.D. J. Inst. Math. Appl. (1977), 20, 485-504 Bivariational Bounds for non-linear problems.

(22) Bourne, D.E. and Kendall, P.C. Vector Analysis and Cartesian Tensors, (2nd Edition). (1977), Nelson

(23) Burrows, B.L. and Perks, A.J. J. Math. Anal. Appl. (1981), 83, 77-86 Variational Principles and Variational Iterative Principles for a class of problems.

(24) Burrows, B.L. and Perks, A.J. J. Phys. A. Math. Gen. 14 (1981), 797-808 Complementary variational principles and variational iterative principles.

(30) Chambers, L.G. Integral Equations (1976). International Textbook Co. Ltd.

(31) Cole, R.J. Proc. Roy. Soc. Edinburgh (1980), 76A, 115-128 Complementary Bivariational Principles for Linear Problems involving Non-self-adjoint operators.

(32) Cole, R.J. and Pack, D.C. Proc. Roy. Soc. (1975), A 347, 239-252 Some Complementary Bivariational Principles for Linear Integral Equations of Fredholm type.

(33) Collins, W.D. Univ. W.S. Math. Res. Center Report No. 1624, (1976), Dual Extremum Principles for Dissipative Systems.

(34) Collins, W.D. Proc. Roy. Soc. Edinburgh (1977), 77A, 273-292 Dual Extremum Principles for the Heat Equation.

(35) Craven, B.D. Mathematical Programming and Control Theory (1978) Chapman and Hall.

(36) Eggleston, H.G. Convexity (1958). Cambridge University Press.

(37) Elsgolc, L.E. Calculus of Variations (1961) Pergamon Press.

(38) Fujita, H. J. Phys. Soc. Japan. (1955), 10 (1), 1-5 Contribution to the theory of upper and lower bounds in boundary value problems.

(39) Gandolfo, G. Economic Dynamics : Methods and Models (1980), North Holland

(40) Gradshteyn, I.S. and Ryzhik, I.H. Table of Integrals, Series and Products. (1980) Academic Press

(41) Jordan, D.W. and Smith, P. Non-linear Ordinary Differential Equations, (1977), Clarendon Press, Oxford

(42) Kreider, B.L. Kuller, R.G. Osberg, D.R. Perkins, F.W. Introduction to Linear Analysis (1966) Addison-Wesley

(43) Kreyszig, E. Introductory Functional Analysis with Applications (1978), John Wiley and Sons.

(44) Mikhlin, S.G. Variational Methods in Mathematical Physics, (1964) Pergamon Press

(45) Milne, R.D. Applied Functional Analysis : An Introductory Treatment (1980), Pitman Publishing Ltd.

(46) Moiseiwitsch, B.Z. Integral Equations (1977), Longman.

(47) Nering, E.D. Linear Algebra and Matrix Theory, 2nd Edition, (1970). Wiley International

(48) Noble, B. Univ. W.S. Math. Res. Cent. Report No. 473 (1964)

(49) Noble, B. Univ. W.S. Math. Res. Center Report No. 643 (1966)

(50) Noble, B. and Sewell, M.J. J. Inst. Math. Appl. (1972), 9, 123-193 On Dual Extremum Principles in Applied Mathematics.

(51) Pogorzelski, W. Integral Equations and their Applications, Vol 1 (1966) Pergamon Press

(52) Rall, L.B. J. Math. Anal. Appl. (1966), 14, 174-184 On Complementary Variational Principles

(53) Robinson, P.D. Non-Linear Functional Analysis and Applications, Ed. L.B. Rall, (1971), 507-576. Academic Press Complementary Variational Principles

(54) Robinson, P.D. and Arthurs, A.M. J. Math. Phys. (1968), 9 (9), 1364-1368 Some variational principles for integral equations.

(55) Robinson, P.D. and Barnsley, M.F. S.I.A.M. J. Numer. Anal. (1979), 16 (1), 135-144 Pointwise Bivariational Bounds on Solutions of Fredholm Integral Equations

(56) Rouch, N. and Mawhin, J. Ordinary Differential Equations. (1980), Pitman Publishing Ltd.

(57) Sewell, M.J. Phil. Trans. Roy. Soc. London (1969), Vol 265, 319-351. On Dual Approximation Principles and Optimisation in Continuum Mechanics.

(58) Sewell, M.J. J. Structural Mech. (1973), 2 (1), 1-32; 2 (2), 135-158
The Governing Equations and Extremum Principles of Elasticity and Plasticity Generated from a Single Functional, Parts I and II.

(59) Sewell, M.J. Duality and Complementarity in Mechanics of Solids, Poloka Akademic Nouk, (1979), Warsaw, 273-349
Degenerate Duality, Catastrophes and Saddle Functionals

(60) Sewell, M.J. Mechanics of Solids, Eds. H.G. Hopkins and M.J. Sewell, (1982), 563-603, Pergamon, Oxford
Legendre Transformations and Extremum Principles

(61) Sloan, D.M. Z. Angew. Math. Phys. (1973), Vol 24. 689-698
Extremum Principles for Magnetohydrodynamic Pipe Flow

(62) Smith, P. Appl. Sci. Res. 24 (1971) 452-466
Some asymptotic extremum principles for Magnetohydrodynamic Pipe Flow

(63) Smith, P. Z. Angew. Math. Phys (1972) Vol 23. 753-764
Some extremum principles for Pipe Flow in Magnetohydrodynamics

(64) Smith, P. J. Inst. Math. Appl. (1976), 18, 129-133
A note on dual extremum principles in Magnetohydrodynamic Pipe Flow

(65) Smith, P. J. Inst. Math. Appl. (1978), 22, 457-465
Approximate Operators and Dual Extremum Principles for Linear Problems

(66) Smith, P. Z. Angew. Math. Mech. (1979), 59, 111-115
On the construction of operators and boundary functionals for dual extremum principles

(67) Smith, P. — Math. Methods Appl. Sci. (1981), 3, 189-199
On the construction of minimising sequences for convex functionals with upper and lower bounding functionals

(68) Smith, P. — Non-linear Analysis, Theory, Methods and Applications (1981), 5 (2), 185-193
Extended Dual Extremum Principles for Non-linear Problems

(69) Smith, P. — Math Nachr. (1983) 113 (In Press)
Some applications of Perturbation Methods to Saddle Point Bounds

(70) Sperb, R. — Maximum Principles and their Applications (1981) Academic Press

(71) Stakgold, I. — Greens Functions and Boundary Value Problems, (1979), Wiley

(72) Swetits, J. and Rogers, C. — J. Math. Anal. Appl. (1978), 62, 445-452
On Complementary Extremum Principles

(73) Timoshenko, S. and Goodier, J.W. — Theory of Elasticity, (1981), 2nd Edition
McGraw-Hill Books Inc.

(74) Vainberg, M.M. — Variational Method and method of monotone operators in the theory of non-linear equations (1973). John Wiley, New York

(75) Walpole, L.J. — J. Inst. Math. Appl. (1974), 14, 113-118
New extremum principles for linear problems

(76) Weinberger, H.F. — A First Course in Partial Differential Equations, (1965) Xerox

(77) Wouk, A. — A Course of Applied Functional Analysis, (1979). John Wiley and Sons.

(67) Smith, P. Math. Methods Appl. Sci. (1981), 3, 189-199
On the construction of minimising sequences for convex functionals with upper and lower bounding functionals

(68) Smith, P. Non-linear Analysis, Theory, Methods and Applications (1981), 5 (2), 185-193
Extended Dual Extremum Principles for Non-linear Problems

(69) Smith, P. Math Nachr. (1983) 113 (In Press)
Some applications of Perturbation Methods to Saddle Point Bounds

(70) Sperb, R. Maximum Principles and their Applications (1981) Academic Press

(71) Stakgold, I. Greens Functions and Boundary Value Problems, (1979), Wiley

(72) Swetits, J. and Rogers, C. J. Math. Anal. Appl. (1978), 62, 445-452
On Complementary Extremum Principles

(73) Timoshenko, S. and Goodier, J.W. Theory of Elasticity, (1981), 2nd Edition McGraw-Hill Books Inc.

(74) Vainberg, M.M. Variational Method and method of monotone operators in the theory of non-linear equations (1973). John Wiley, New York

(75) Walpole, L.J. J. Inst. Math. Appl. (1974), 14, 113-118
New extremum principles for linear problems

(76) Weinberger, H.F. A First Course in Partial Differential Equations, (1965) Xerox

(77) Wouk, A. A Course of Applied Functional Analysis, (1979). John Wiley and Sons.